Thèse délivrée par

Université de Lille et Ghent University

Pour l'obtention du titre de Docteure en Sciences Économiques

présentée et soutenue publiquement par

Juliana Sanchez Ramirez

Le 30 Janvier 2026

Applications, défis et mises en œuvre de l'IA dans les entreprises B2B

Membres du Jury:

Directeurs de thèse:	Dr. Kristof Coussement, Professor, IÉSEG School of Management
	Dr. Dries F. Benoit, Professor, Ghent University
	Dr. Arno De Caigny, Professor, IÉSEG School of Management
Président du jury:	Dr. Dominique Crié, Professor, Université de Lille
Rapporteurs:	Dr. Francisco Villarroel Ordenes, Professor, University of Bologna
	Dr. Sebastián Maldonado Alarcón, Professor, Universidad de Chile
Examinateurs:	Dr. Nadia Steils, Professor, HEC Liège
	Dr. Matthias Bogaert, Professor, Ghent University
Invité:	Mr. Tim Bernaerdt, Director of Enfocus

L'université de Lille n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

LABORATOIRE DE RATTACHEMENT :
Lille Economie Management (LEM - UMR CNRS 9221)Laboratoire de recherche rattaché à l'IAE de Lille et à la Fédération Universitaire Polytechnique de Lille (FUPL).
Préparation de la thèse sur le site de l'IESEG School of Management, 3 Rue de la Digue, 59000 Lille.

Acknowledgements

This PhD was a particularly multi-stakeholder process, and I am sincerely grateful to everyone whose involvement, feedback, and collaboration contributed to its completion. Without their contributions, this PhD would not have been possible. I would also like to express my sincere thanks to the jury members for their willingness to take part in this process and for their readiness to engage with my work.

First, I would like to express my sincere appreciation to IESEG School of management, Enfocus, and the region Hauts de France whose trust and financial support enabled me to pursue this research.

My deepest gratitude and appreciation to my thesis supervisors Kristof, Arno and Dries for their invaluable trust and valuable feedback. Their expertise, dedication, and support have been fundamental to the successful completion of this PhD and to my professional and personal growth. I am profoundly grateful for the opportunity to learn from their experience, pragmatic approach, and high standards of academic excellence. Their patience, openness, and disposition taught me to be more confident in my own judgment, to question my assumptions thoughtfully, and to stay grounded and resilient when facing obstacles.

Additionally, I am immensely grateful for the collaboration and support from Lauren. Her generosity, openness, and willingness to share her time and expertise have been truly invaluable throughout this journey. I deeply appreciate her thoughtful insights, her readiness to help whenever challenges arose, and the genuine care she brought into our work together. Her presence has made this path more enriching and inspiring.

Moreover, I am grateful to the Enfocus team for their openness to new ideas, their generous access to rich data resources, and their genuine curiosity for exploring novel applications. Their willingness to involve me in diverse activities provided valuable opportunities to learn about the industry and to understand the practical context of my research more deeply. I would like to express my special thanks to Tim for his sustained interest in my work and for taking the time to follow up on the projects, even with his demanding schedule. I truly appreciate his engagement and encouragement.

I am also thankful to the colleagues from Ghent University who generously welcomed me. Always kind and open to discussion, willing to provide thoughtful and relevant feedback, which greatly enriched my time there. My best wishes for all your future endeavors. I am equally grateful to all my colleagues at IESEG, Minh, Khaoula, Stephanie, Philipp, and Emil, whose experience, guidance, and thoughtful advice led me throughout these three years, even as our paths eventually diverged. I would like to give a special thanks to Emil, with whom I had the pleasure of working closely. His practical tips and amazing stories made this journey genuinely enjoyable. I am also grateful to the colleagues who joined IESEG alongside me or even after, for the camaraderie, encouragement, and everyday support we shared. Special thanks to Pele, Umur, Song, Divya, Fernando and Alice, their presence brought motivation and a sense of family that made the demanding periods easier to navigate. I could not have asked for better colleagues. Thank you for all the great memories, I wish you all the best in what lies ahead.

To my closest friends, David, Daniel, Sara, Juan, Angela, Ana, Samara, Juanjo and Pedro who regardless the time and distance, have always been there for me, believing in me and supporting

my decisions with unwavering trust and affection. To Gauthier, whose company and never-ending encouragement have been essential to follow this journey. Your kindness, patience, and steady presence helped me navigate difficult moments and appreciate the meaningful ones. Thank you for being a constant source of warmth, understanding, and unwavering support.

Special thanks to my family, my parents and my sister, for your love, motivation, inspiration, and trust. Your constant presence, even from afar, has been a source of strength and joy throughout these years. Thank you for your unwavering belief in me and for the warmth, stability, and encouragement you continue to bring into my life.

Juliana Sanchez Ramirez
December 2025

Doctoral committee

Directeurs de thèse:
Dr. Kristof Coussement, Professor, IÉSEG School of Management
Dr. Dries F. Benoit, Professor, Ghent University
Dr. Arno De Caigny, Professor, IÉSEG School of Management

Président du jury:

Dr. Dominique Crié, Professor, Université de Lille

Rapporteurs:

Dr. Francisco Villaroel Ordenes, Professor, University of Bologna
Dr. Sebastián Maldonado Alarcón, Professor, Universidad de Chile

Examinateurs:

Dr. Nadia Steils, Professor, HEC Liège
Dr. Matthias Bogaert, Professor, Ghent University

Invité:

Mr. Tim Bernaerdt, Director of Enfocus

List of publications

This dissertation is based on three individual studies

1. **Sanchez Ramirez, J.**, Coussement, K., De Caigny, A., Benoit F., Guliyev E. (2024). Incorporating usage data for B2B churn prediction modeling. *Industrial Marketing Management*. Received 31 August 2023, Accepted 23 May 2024, Published: July 2024.

   Preliminary results were also presented at:

   - 37th Annual Conference of the Belgian Operational Research Society (ORBEL 2023), May 2023. Liège, Belgium.
   - 23rd International Conference of the International Federation of Operational Research Societies (IFORS 2023), July 2023. Santiago, Chile.

2. **Sanchez Ramirez, J.**, Coussement, K., De Caigny, A., Benoit F., D. Dynamic Predictive Modeling of SaaS Trial Conversions: Evidence from a B2B setting. Working paper to be submitted to *Production and Operations Management*.

   Preliminary results were also presented at:

   - 38th Annual Conference of the Belgian Operational Research Society (ORBEL 2024), Feb 2024. Antwerp, Belgium.
   - 46th INFORMS Society for Marketing Science Conference (ISMS 2024), June 2024. Sydney, Australia.

3. **Sanchez Ramirez, J.**, Waardenburg, L., Coussement, K., De Caigny, A., Benoit F., D. Mirroring in AI Design: Challenges and Organizational Impacts. Working paper to be submitted to Information and Organization.

   Preliminary results were also presented at:

   - 85th Annual Meeting of the Academy of Management (AOM 2025), July 2025. Copenhagen, Denmark.

Contents

Abstract

Résumé Général

Cette thèse examine l'intégration de l'intelligence artificielle (IA) dans les environnements logiciels business-to-business (B2B) au moyen d'un dispositif de recherche multi-méthodes mené au sein d'une entreprise européenne de logiciels B2B. Motivée par le rôle stratégique croissant de l'IA dans la transformation des processus décisionnels fondés sur les données et des modes de coordination organisationnelle en contexte B2B, cette recherche analyse comment les applications d'IA créent de la valeur, rencontrent des contraintes analytiques et organisationnelles, et s'intègrent dans les pratiques des entreprises. Pour structurer cette investigation, la thèse développe un cadre analytique composé de trois dimensions interreliées : les applications, les défis et la mise en œuvre.

Ces trois dimensions sont examinées à travers une combinaison d'études quantitatives et qualitatives réalisées dans un environnement logiciel B2B. Dans la dimension des applications, la recherche montre comment les données comportementales et interactionnelles peuvent être transformées en résultats analytiques éclairant les décisions d'acquisition et de rétention des clients. Les études démontrent que les schémas d'usage, les trajectoires d'engagement et les comportements d'essai peuvent être modélisés de manière à élargir la base informationnelle de l'action managériale. Dans la dimension des défis, les analyses révèlent les conditions analytiques et organisationnelles qui influencent la faisabilité et l'utilité des applications d'IA. Celles-ci incluent l'effort nécessaire pour structurer des données hétérogènes, la nécessité de préserver l'interprétabilité lorsque les modèles deviennent plus complexes, ainsi que les contraintes générées par les structures de coordination et de gouvernance existantes. La dimension de mise en œuvre retrace le processus organisationnel par lequel les outils analytiques sont conçus, développés et intégrés dans les pratiques. Les résultats montrent que ce processus se déploie en phases successives de définition des données, de spécification du modèle et d'intégration de l'outil, façonnées par les interactions entre experts techniques, managers et parties prenantes organisationnelles.

Ensemble, les trois études offrent une base empirique intégrée permettant de comprendre comment l'IA s'ancre au sein des organisations B2B. Les analyses montrent que les applications fondées sur les données évoluent à travers des processus qui transforment les informations comportementales en connaissances actionnables, permettant aux entreprises d'anticiper les besoins des clients et de soutenir des décisions éclairées. Les résultats démontrent également que les contraintes techniques et organisationnelles façonnent conjointement la faisabilité et l'impact des initiatives d'IA, la réussite de leur mise en œuvre dépendant d'une articulation durable entre infrastructures de données, modèles analytiques, rôles managériaux et routines opérationnelles. En synthétisant ces éléments, la thèse propose une perspective multi-niveaux sur l'intégration de l'IA en contexte B2B, montrant que l'IA contribue à la performance des entreprises par trois mécanismes récursifs — création de valeur, préservation de la valeur et institutionnalisation de la valeur — qui se déploient à différents stades du cycle de vie client et organisationnel. Pris ensemble, ces mécanismes montrent que la valeur des intégrations d'IA en contexte B2B émerge de l'alignement et de la co-évolution continus des actifs de données, des configurations analytiques et des pratiques organisa-

tionnelles, grâce auxquels les systèmes prédictifs deviennent à la fois opérationnellement pertinents et institutionnellement pérennes.

Mots-clés: Intelligence artificielle, Business-to-Business (B2B), Apprentissage automatique, Logiciel en tant que service (SaaS), Données d'usage.

General Abstract

This dissertation examines the integration of artificial intelligence (AI) in business-to-business (B2B) software environments through a multi-method research design conducted within a B2B European software company. Motivated by the growing strategic role of AI in reshaping data-driven decision processes and organizational coordination in B2B environments, this research investigates how AI applications create business value, encounter analytical and organizational constraints, and become embedded in organizational practice. To guide this purpose, the dissertation develops an analytical framework comprising three interrelated dimensions, applications, challenges, and implementation.

The three dimensions are examined through a combination of quantitative and qualitative studies conducted within a B2B software environment. Within the applications dimension, the research shows how behavioral and interaction data can be transformed into analytical outputs that inform customer acquisition and retention decisions. The studies demonstrate how usage patterns, engagement trajectories, and trial usage can be modeled in ways that extend the informational basis of managerial action. In the challenges dimension, the analyses reveal both analytical and organizational conditions that shape the feasibility and usefulness of AI applications. These include the effort required to structure heterogeneous data, the need to maintain interpretability when working with more complex analytical designs, and the constraints created by existing coordination and governance structures. The implementation dimension traces the organizational process through which analytical tools are conceived, designed, and brought into use. The findings show how implementation unfolds through successive phases of data definition, model specification, and tool integration, shaped by the interactions between technical experts, managers, and relevant organizational stakeholders.

Together, the three studies offer an integrated empirical foundation for understanding how AI becomes embedded within B2B organizations. The analyses show that data-driven applications evolve through processes that translate behavioral information into actionable insights, enabling firms to anticipate customer needs and support informed decision-making. The findings further demonstrate that technical and organizational constraints together shape the feasibility and impact of AI initiatives, as successful implementation depends on sustained alignment among data infrastructures, analytical models, managerial roles, and operational routines. Synthesizing these insights, the dissertation advances a multi-level perspective on AI integration in B2B contexts, demonstrating that AI contributes to business performance through three recursive mechanisms, value creation, value preservation, and value institutionalization, that unfold across different stages of the customer and organizational lifecycle. Taken together, these mechanisms show that the value of AI integrations in B2B contexts emerges from the ongoing alignment and co-evolution of data assets, analytical configurations, and organizational practices, through which predictive systems become operationally relevant and institutionally sustained.

Keywords: Artificial Intelligence, Business-to-Business (B2B), Machine Learning, Software-as-a-Service (SaaS), Usage Data.

Chapter 1

Introduction

Chapter 1
Introduction

1.1 Introduction Générale

1.1.1 Contexte et motivation

L'environnement commercial contemporain se caractérise par une prolifération sans précédent de sources de données numériques, créant des écosystèmes informationnels complexes dans lesquels les organisations sont confrontées à des flux continus de données opérationnelles, transactionnelles et interactionnelles (Mcafee and Brynjolfsson, 2012; Grover et al., 2018; Kellogg et al., 2020). La croissance qui en résulte en termes de volume, de vélocité et de variété des données a déplacé le défi stratégique de l'acquisition des données vers l'extraction d'informations exploitables à partir de jeux de données hétérogènes et de haute dimensionnalité, qui dépassent les capacités cognitives humaines et les méthodes analytiques traditionnelles (Kellogg et al., 2020; George et al., 2014).

Cette complexité met en évidence les limites des approches analytiques conventionnelles. Les méthodes statistiques traditionnelles et les systèmes d'aide à la décision fondés sur des règles, conçus pour des données structurées et des schémas stables, ne peuvent pas traiter efficacement les flux informationnels actuels. Conscientes de cet écart, les organisations se tournent de plus en plus vers l'intelligence artificielle (IA) comme solution transformatrice. Les avancées récentes en apprentissage automatique (ML), apprentissage profond (DL) et traitement automatique du langage naturel (NLP) ont démontré des capacités remarquables, allant au-delà des performances humaines en reconnaissance d'images jusqu'à la génération de modèles linguistiques sophistiqués (Jordan and Mitchell, 2015; Lecun et al., 2015; Krizhevsky et al., 2017). Ces percées techniques ont catalysé une large intégration de l'IA dans les industries, les entreprises nativement digitales et les applications orientées consommateurs menant cette transformation (Mcafee and Brynjolfsson, 2012).

L'IA a déjà généré une valeur substantielle dans divers secteurs. Les entreprises de plateformes ont capturé des milliards de revenus grâce aux systèmes de recommandation et à la publicité ciblée (Gomez-Uribe and Hunt, 2015; Covington et al., 2016; Linden et al., 2017). Les institutions financières ont automatisé des processus complexes avec des retours mesurables (Bahoo et al., 2024), tandis que les acteurs du commerce de détail utilisent l'analytique prédictive pour optimiser leurs opérations et l'expérience client (Fildes et al., 2022). Ces exemples illustrent le potentiel transformateur de l'IA pour les entreprises, bien que l'ampleur des bénéfices réalisés dépende de conditions contextuelles et d'objectifs spécifiques (Ledro et al., 2025).

Les contextes business-to-business (B2B) représentent un environnement dans lequel ces conditions deviennent particulièrement marquées. Dans les marchés B2B, les relations ne sont pas de simples transactions mais des échanges à long terme fondés sur la confiance, nécessitant un engagement et une collaboration continus (Morgan and Hunt, 1994). Les entreprises B2B reposent souvent sur des réseaux décisionnels complexes et des liens professionnels qui dépassent les frontières contractuelles formelles (Mojir and Anbil, 2025).

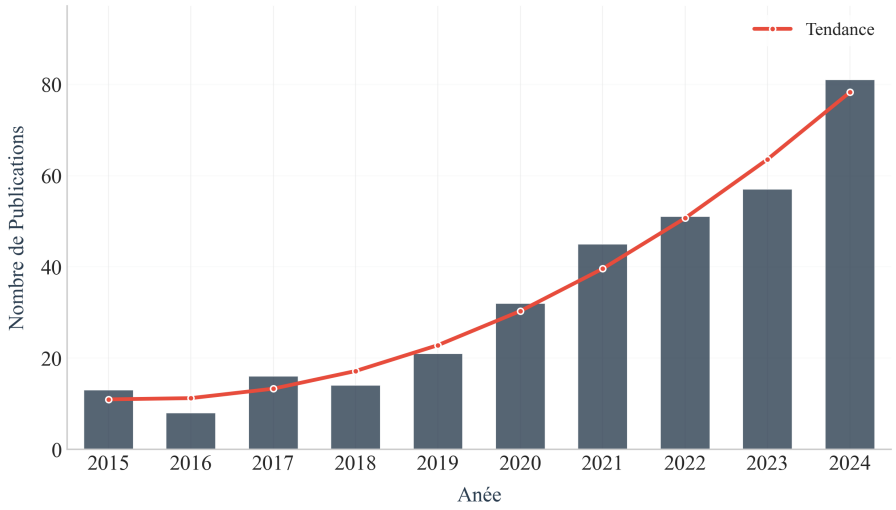


Figure 1.1: Distribution temporelle des publications sur l'IA en B2B

La nature relationnelle des environnements B2B a des implications directes sur la manière dont l'IA peut y être intégrée. Les cycles de vente prolongés réduisent la fréquence des résultats observables, limitant ainsi la portée de la validation empirique des informations générées par l'IA (Moradi and Dass, 2022; Syam and Sharma, 2018). Les processus décisionnels impliquant de multiples parties prenantes introduisent une ambiguïté dans l'attribution de la valeur, puisque les sorties prédictives sont interprétées et utilisées par divers acteurs à différents niveaux industriels (Purmonen et al., 2023). De plus, l'efficacité des systèmes décisionnels fondés sur l'IA dépend de la disponibilité, de la structuration et de l'intégration de données souvent limitées et dépendantes des relations, imposant des exigences supplémentaires en matière de spécification et d'évaluation des modèles dans les environnements B2B (Lim et al., 2019).

Pris ensemble, ces éléments positionnent les environnements B2B comme un contexte critique pour examiner l'intégration de l'IA, où les systèmes prédictifs doivent opérer au sein de réseaux de confiance, d'interactions soutenues et d'interdépendances stratégiques. Dans ces contextes dépendants des relations, l'intégration de l'IA soulève en outre des questions fondamentales concernant la disponibilité des données et les délais de rétroaction, l'attribution de la valeur entre plusieurs unités décisionnelles, l'interopérabilité des nouveaux systèmes avec les infrastructures existantes, ainsi que la gouvernance de l'allocation des rôles entre humains et IA (Nyadzayo et al., 2020). Ensemble, ces complexités relationnelles et techniques soulignent l'importance de développer des approches analytiques et des systèmes prédictifs capables de s'adapter aux dynamiques des interactions B2B, tout en tenant compte des contraintes organisationnelles et informationnelles, et en soutenant une adoption responsable de l'IA préservant les relations.

Afin de situer ce travail dans le paysage scientifique plus large, une revue structurée des publications évaluées par les pairs portant sur l'IA dans les contextes B2B entre 2015 et 2024 a été réalisée. La littérature révèle une trajectoire de croissance claire et accélérée, les contributions augmentant substantiellement après 2019 et l'IA en B2B passant d'un sujet marginal à un domaine de recherche reconnu. Comme l'illustre la Figure 1.1, l'activité de publication s'est accrue de manière régulière depuis 2019, reflétant l'émergence de l'IA en B2B comme domaine académique établi. Au fil du temps, l'accent thématique s'est également déplacé des premières applications fonctionnelles — telles que la prévision des ventes et la prédiction de la demande — vers des perspectives davantage centrées sur la technologie et, plus récemment, vers des méthodes analytiques diversifiées et des domaines stratégiques. Cette évolution, résumée dans la Figure 1.2, illustre la transition entre des études à orientation opérationnelle et des travaux mobilisant des techniques d'IA avancées et abordant des questions managériales plus larges.

Malgré cette croissance, la littérature existante présente plusieurs limites. De nombreuses contributions

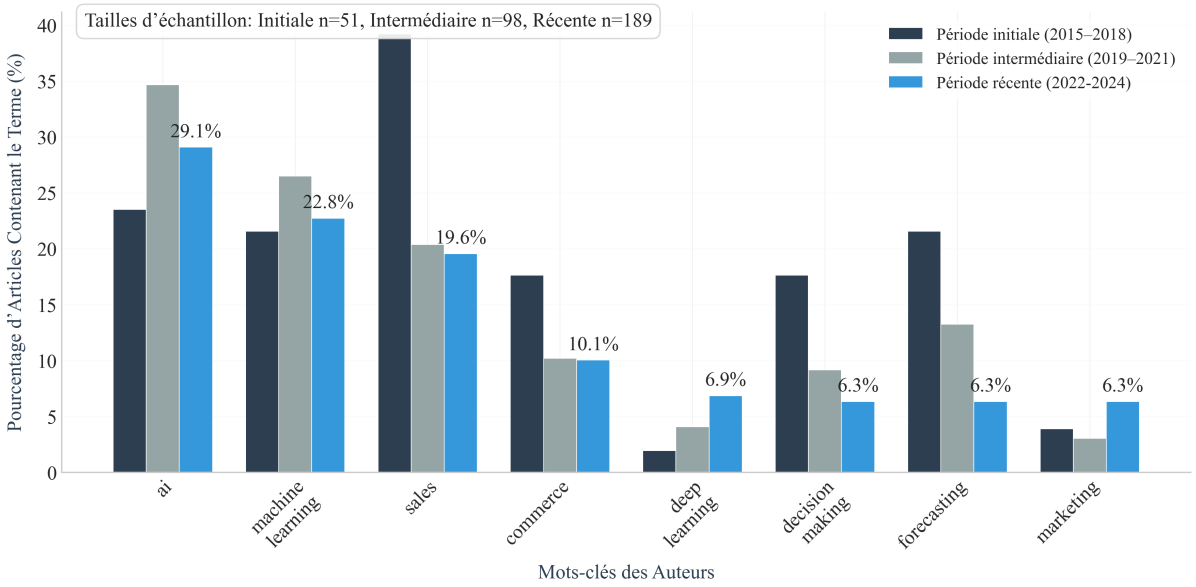


Figure 1.2: Évolution temporelle des mots-clés des auteurs selon les périodes de recherche

illustrent des cas d'usage dans des domaines tels que la prévision des ventes, la gestion de la clientèle et l'analytique marketing (Gordini and Veglio, 2017; Jahromi et al., 2014; Gattermann-Itschert and Thonemann, 2022). Des travaux récents ont élargi cette perspective en reliant l'IA au cycle de vie du client et en documentant les perceptions industrielles des bénéfices, obstacles et opportunités (Moradi and Dass, 2022). Ces études établissent le potentiel de l'IA en B2B et démontrent des voies claires vers une efficacité accrue et une meilleure gestion de la relation client. Toutefois, elles demeurent concentrées sur un ensemble restreint de domaines fonctionnels et s'appuient souvent sur des sources de données similaires, négligeant la manière dont de nouveaux types de données, combinés à des techniques avancées comme l'apprentissage profond, peuvent transformer la création de valeur dans les environnements B2B. L'écart réside ainsi dans l'élargissement des preuves empiriques vers des applications pratiques exploitant des sources de données nouvelles et reliant explicitement les modèles d'IA à des résultats commerciaux mesurables, afin d'élargir notre compréhension de la contribution de l'IA à la performance B2B.

Une deuxième limite concerne la fragmentation des recherches portant sur les contraintes à l'intégration de l'IA. Bien que des travaux antérieurs aient identifié des contraintes techniques et organisationnelles, ils manquent d'une perspective intégrée. Les défis techniques, tels que la qualité, l'accès et la robustesse des données (Dwivedi and Wang, 2022), apparaissent souvent déconnectés des obstacles organisationnels, notamment le manque de compétences, la résistance culturelle et les restrictions de ressources. Dans le marketing B2B en particulier, les coûts élevés des systèmes avancés d'IA et la difficulté de les aligner sur des jeux de données hétérogènes et des structures décisionnelles multi-acteurs constituent des obstacles supplémentaires (Chatterjee et al., 2024). Cependant, ces barrières sont fréquemment considérées isolément, sans les relier aux configurations techniques des systèmes d'IA ou à la valeur commerciale qu'ils visent à produire.

Une troisième limite porte sur les processus organisationnels façonnant la mise en œuvre de l'IA. Une grande partie de la littérature existante met l'accent sur des démonstrations de preuve de concept ou des applications pilotes, offrant une compréhension limitée de la manière dont les systèmes d'IA sont adaptés, gouvernés et institutionnalisés au sein des entreprises. Bien que des travaux antérieurs mettent en lumière le potentiel de l'IA pour transformer les ventes, le marketing, l'engagement client et la collaboration interorganisationnelle (Paschen et al., 2019), et que des études récentes reconnaissent les défis d'alignement et d'intégration (Keegan et al., 2024), les analyses systématiques portant sur la

manière dont l'IA interagit avec des cycles de vente étendus et des processus décisionnels multi-acteurs demeurent rares. Développer une telle compréhension est essentiel pour expliquer comment l'IA évolue de l'expérimentation initiale vers une utilisation pérenne dans les contextes B2B.

Pris ensemble, ces écarts soulignent la nécessité d'une compréhension plus intégrée de la manière dont l'IA peut créer de la valeur, de la façon dont les choix méthodologiques influencent la performance prédictive, et de la manière dont les conditions organisationnelles déterminent le succès de l'adoption de l'IA dans les contextes B2B. L'objectif de cette thèse est de répondre à ces limitations en examinant la conception, l'évaluation et l'intégration organisationnelle de systèmes prédictifs dans des environnements B2B fortement relationnels. Pour faire avancer cette compréhension, la dissertation s'articule autour de trois questions de recherche. Premièrement, comment différentes méthodes d'IA et de configurations de données peuvent-elles être conçues et évaluées afin de générer une valeur prédictive et opérationnelle dans les environnements B2B ? Deuxièmement, quelles conditions techniques et organisationnelles limitent l'évolutivité, la fiabilité et la réalisation de la valeur des applications d'IA dans les contextes B2B ? Troisièmement, comment les applications d'IA deviennent-elles intégrées, adaptées et institutionnalisées au sein des organisations B2B au fil du temps, et comment ces processus d'implémentation interagissent-ils avec les défis émergents et les configurations analytiques ? Les chapitres qui suivent abordent ces questions à travers une combinaison d'analyses quantitatives et qualitatives, offrant une compréhension multi-niveaux de l'évolution de l'IA, depuis les premières applications jusqu'à son intégration durable dans les environnements B2B.

1.1.2 Cadre analytique

En s'appuyant sur ces questions de recherche, la thèse introduit un cadre analytique qui conceptualise l'intégration de l'IA dans les environnements B2B comme un processus continu se déployant selon trois dimensions interconnectées : les Applications, les Défis et l'Implémentation. Ce cadre structure l'analyse nécessaire pour examiner comment les applications des systèmes prédictifs sont conçues et évaluées (RQ1), comment les défis techniques et organisationnels façonnent leur évolutivité et la réalisation de leur valeur (RQ2), et comment les outils d'IA deviennent intégrés, adaptés et institutionnalisés au sein des entreprises (RQ3). L'organisation de la thèse autour de ces dimensions fournit une base cohérente pour analyser les processus technologiques, analytiques et organisationnels à travers lesquels l'IA évolue et crée de la valeur dans des environnements B2B à forte intensité relationnelle.

Le cadre s'appuie sur des travaux empiriques et théoriques antérieurs afin de saisir l'interaction entre mécanismes technologiques et dynamiques organisationnelles. Chaque dimension offre un prisme analytique distinct et, ensemble, elles soutiennent la conception de recherche multiniveau de la dissertation, dans laquelle les analyses quantitatives évaluent la performance des modèles et les implications de différentes configurations de données, tandis que les analyses qualitatives les contextualisent et les interprètent dans la pratique organisationnelle.

La dimension *applications* examine comment les techniques d'IA génèrent de la valeur commerciale lorsqu'elles sont alignées avec les tâches et les données caractéristiques du B2B. La recherche antérieure suggère que l'impact de l'IA dépend de l'adéquation entre la méthode, la structure des données et le contexte décisionnel. Par exemple, Syam and Sharma (2018) montrent comment l'IA améliore la gestion des ventes B2B via le lead scoring dynamique et la qualification des opportunités en intégrant des prédictions dans les processus de travail des commerciaux. De même, Chatterjee et al. (2022) constatent que les systèmes CRM reposant sur l'IA dans les marchés industriels sont associés à une plus grande satisfaction relationnelle et à une performance organisationnelle améliorée, soulignant l'importance d'intégrer les sorties analytiques dans les routines orientées client. En élargissant ces études spécifiques, Vlačić et al. (2021) réalisent une revue systématique des applications de l'IA en marketing et montrent

que l'apprentissage automatique et l'apprentissage profond étendent les capacités analytiques en traitant de grands jeux de données hétérogènes et en identifiant des schémas comportementaux excédant les limites des modèles statistiques traditionnels. L'étude soutient que ces techniques ne créent de la valeur que lorsqu'elles sont alignées avec la disponibilité des données, la préparation organisationnelle et l'interprétabilité managériale, conditions qui déterminent si les méthodes analytiques avancées peuvent devenir actionnables au sein des entreprises.

Dans ce cadre, les *applications* représentent les mécanismes technologiques à travers lesquels l'IA transforme les opérations organisationnelles et crée de la valeur économique (Enholm et al., 2022). Elles offrent un prisme analytique permettant d'examiner comment différentes méthodologies d'IA interagissent avec les environnements de données propres aux organisations B2B. Les analyses quantitatives évaluent la manière dont les architectures de modèles, les processus d'entraînement et les nouvelles sources de données transforment la prédiction et le soutien à la décision, tandis que les approches qualitatives montrent comment les sorties analytiques sont définies, appropriées, intégrées et institutionnalisées dans la pratique organisationnelle.

La deuxième dimension, les *défis*, analyse les conditions qui limitent l'intégration de l'IA et empêchent la réalisation cohérente de valeur économique. Ces défis proviennent à la fois de sources techniques et organisationnelles qui déterminent la mesure dans laquelle les capacités analytiques peuvent être converties en résultats opérationnels. D'un point de vue méthodologique, la recherche souligne comment la prolifération de données non structurées et hétérogènes a profondément transformé le paysage analytique, stimulant le développement d'algorithmes avancés et de techniques de représentation des connaissances pour gérer cette complexité (Agarwal and Dhar, 2014). Pourtant, comme le soulignent Mikalef and Gupta (2021), dans les contextes B2B, la disponibilité limitée de données de grande ampleur et de haute granularité restreint le développement de modèles robustes et généralisables, réduisant ainsi la capacité des entreprises à convertir le potentiel analytique de l'IA en création de valeur durable.

Au niveau organisationnel, les études mettent en évidence des limitations en matière de gouvernance, de coordination et de capacités, qui freinent l'évolutivité des initiatives d'IA. Dans ce cadre, Weber et al. (2023) identifient la collaboration interfonctionnelle comme un défi organisationnel critique, notant que l'opacité inhérente aux systèmes d'IA amplifie le fossé informationnel entre les départements techniques et les autres fonctions de l'entreprise. De même, Schneider et al. (2023) soulignent que les organisations rencontrent des difficultés persistantes à établir une gouvernance robuste des données, à développer des capacités analytiques suffisantes et à assurer la coordination entre les unités fonctionnelles, autant de facteurs qui limitent l'échelle et la réalisation de la valeur générée par l'IA. Dans le marketing B2B, Keegan et al. (2024) montrent que l'intégration de l'IA est souvent entravée par des contradictions fondamentales entre les fournisseurs de solutions et les entreprises clientes. Ces contradictions se manifestent à travers des valeurs organisationnelles non alignées et une absence de compréhension commune de la manière dont l'IA peut créer de la valeur, des perturbations des pratiques établies lorsque les technologies d'IA sont assimilées, et des divergences entre les bénéfices attendus et réalisés parmi les parties prenantes managériales.

Dans ce cadre, la dimension *défis* offre un prisme conceptuel pour analyser les conditions dans lesquelles l'intégration de l'IA dans les contextes B2B ne parvient pas à produire de manière cohérente ou scalable de la valeur. La recherche montre de plus en plus que la progression entre potentiel analytique et valeur réalisée est non linéaire, façonnée par l'interdépendance entre systèmes techniques et systèmes organisationnels. La dimension *défis* saisit ces points de friction, où les désalignements entre rigueur méthodologique, infrastructures de données et capacités organisationnelles perturbent l'évolutivité et la fiabilité des résultats de l'IA. Les approches quantitatives révèlent les vulnérabilités méthodologiques qui limitent la robustesse et la généralisabilité des modèles, tandis que les analyses qualitatives mettent au jour des tensions de gouvernance, de coordination et d'interprétation qui freinent l'intégration de l'IA

dans la pratique organisationnelle. Ensemble, ces perspectives renforcent une compréhension intégrée de l'IA comme système sociotechnique, dans lequel les défis constituent des éléments inhérents définissant les frontières d'une création de valeur durable dans les contextes B2B.

La troisième dimension, les *implémentations*, concerne les processus par lesquels les applications d'IA sont adoptées et institutionnalisées au sein des organisations. À mesure que les entreprises intègrent l'IA dans leurs processus décisionnels, de nouveaux schémas de collaboration, de responsabilité et de contrôle émergent (Von Krogh, 2018). La recherche montre également que la mise en œuvre efficace dépend du développement de capacités organisationnelles complémentaires facilitant l'intégration au-delà des frontières internes de l'entreprise. Dans les contextes B2B, Dubey et al. (2021) soulignent l'importance des capacités de gestion d'alliances, qui permettent aux organisations de coordonner les partenariats et de partager des ressources stratégiques essentielles à la pleine réalisation du potentiel de l'IA.

Dans ce cadre, la dimension *implémentations* saisit l'évolution de l'interaction entre systèmes technologiques et structures organisationnelles à mesure que l'IA devient intégrée dans les processus métiers. Elle s'intéresse à la manière dont la mise en œuvre se déroule à travers des cycles d'adaptation, d'apprentissage et de coordination qui redéfinissent les rôles, les flux de travail et les mécanismes de gouvernance. Une perspective technique permet d'évaluer comment l'IA influence la performance et l'efficacité opérationnelle, tandis qu'une perspective organisationnelle met en lumière l'évolution des routines, des pratiques décisionnelles et des arrangements de gouvernance qui se forment autour des nouvelles capacités technologiques. Ensemble, ces perspectives éclairent les dynamiques par lesquelles l'IA dépasse sa fonction technique pour devenir une force organisationnelle intégrée dans les environnements B2B.

Bien que distinctes sur le plan analytique, les trois dimensions — Applications, Défis et Implémentation — sont empiriquement imbriquées et mutuellement renforçantes. Les travaux antérieurs ont généralement examiné ces domaines séparément (performance algorithmique, obstacles organisationnels ou processus d'adoption), mais des preuves croissantes indiquent qu'ils coévoluent dans un processus récursif d'intégration de l'IA (George et al., 2014; Mariani et al., 2023). Dans les environnements B2B, où les données sont distribuées à travers les frontières organisationnelles et où les décisions impliquent plusieurs acteurs, ce caractère récursif devient particulièrement prononcé. Le déploiement de nouvelles applications d'IA engendre souvent des défis inattendus, tels que des dépendances accrues aux données générées par les clients, la nécessité d'interprétabilité dans les processus de vente ou de service complexes, et l'émergence de tensions de gouvernance concernant la propriété des données et la responsabilité (Moradi and Dass, 2022). Les efforts visant à résoudre ces problèmes — par l'amélioration des infrastructures de données, une plus grande transparence des modèles et le développement de capacités analytiques interfonctionnelles — créent les conditions permettant de nouvelles applications et un affinement continu, renforçant un cycle soutenu d'adaptation technologique et organisationnelle.

À mesure que les systèmes d'IA sont intégrés, ils transforment les environnements informationnels et organisationnels qui les contraignaient initialement. Dans les contextes B2B, ces effets de rétroaction sont amplifiés par des cycles de vente prolongés, des dépendances de service et des mécanismes de coordination interentreprises. Ce schéma récursif est également reflété dans les fondements méthodologiques de la recherche en IA. Les études sur les cycles de vie du machine learning soulignent que la fiabilité des modèles et la réalisation de valeur dépendent de raffinements itératifs et de retours contextuels plutôt que de conceptions statiques (Sculley et al., 2015). Le maintien de la performance nécessite des ajustements continus face à des données hétérogènes, l'évolution des portefeuilles clients et la transformation des objectifs organisationnels. La recherche sur l'IA renforce cet alignement en soulignant que les progrès dépendent de plus en plus de l'amélioration de la qualité, de la représentativité et de la gouvernance des données, plutôt que de la complexification des architectures de modèles.

Comme illustré dans la Figure 1.3, les trois dimensions ne sont pas traitées comme des catégories

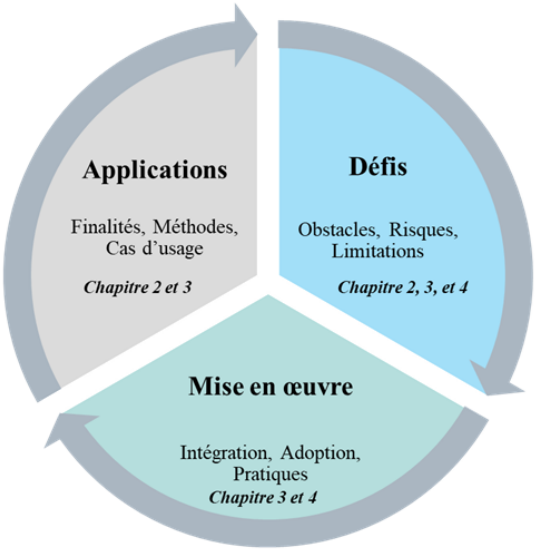


Figure 1.3: Cadre conceptuel : interactions entre applications, défis et implantation

séparées, mais comme des éléments interdépendants d'un processus continu. Le cadre conceptualise l'intégration de l'IA dans les contextes B2B comme un processus cumulatif et évolutif. Il fournit une base pour analyser comment les applications, les contraintes techniques et organisationnelles, et les processus de mise en œuvre se développent et se façonnent mutuellement au fil du temps. Méthodologiquement, la conception de recherche multiniveau renforce cette perspective récursive en reliant les approches quantitatives et qualitatives au sein d'une logique analytique unifiée. Ensemble, ces couches permettent une compréhension multiniveau de l'intégration de l'IA, retraçant comment les capacités technologiques, la complexité analytique et la pratique organisationnelle coévoluent dans la création et la stabilisation de valeur économique dans les environnements B2B.

En somme, en conceptualisant ces trois dimensions comme interconnectées plutôt que séquentielles, le cadre souligne que l'intégration efficace de l'IA dans les environnements B2B ne dépend pas d'un facteur isolé, mais des relations synergétiques entre eux. Il invite chercheurs et praticiens à dépasser les analyses compartimentées pour adopter une perspective intégrée sur la manière dont les systèmes d'IA évoluent en tant qu'éléments constitutifs du tissu organisationnel et interorganisationnel. Ce faisant, il propose une compréhension plus riche et plus opérationnelle des dynamiques sociotechniques et méthodologiques à travers lesquelles la promesse de l'IA est réalisée, adaptée et soutenue dans les environnements B2B complexes.

Afin de démontrer comment le cadre conceptuel structure l'analyse empirique, les trois dimensions sont opérationnalisées dans l'ensemble de la thèse de manière chevauchante mais structurée. La dimension *Applications* est examinée dans les Chapitres 2 et 3, qui analysent comment différentes méthodes d'IA et configurations de données génèrent de la valeur dans des contextes B2B de rétention et d'acquisition. La dimension *Défis* s'étend des Chapitres 2 à 4, puisque des contraintes analytiques, opérationnelles et organisationnelles émergent tout au long du développement, de l'évaluation et du déploiement des systèmes d'IA. La dimension *Implémentation* est principalement traitée dans le Chapitre 4, avec des éclairages pratiques supplémentaires issus du Chapitre 3, en analysant comment les applications d'IA sont adoptées, adaptées et pérennisées au sein des processus organisationnels. Ensemble, ces chapitres opérationnalisent collectivement le cadre, assurant la cohérence entre le modèle conceptuel et les contributions empiriques de la thèse.

1.1.3 Approche méthodologique

La thèse adopte une stratégie de recherche multiméthode qui intègre l'expérimentation quantitative en apprentissage automatique et l'analyse qualitative. Cette conception permet d'étudier les systèmes d'IA selon des perspectives complémentaires : comment la valeur prédictive est générée à partir des données et des modèles, comment des contraintes émergent lors de l'intégration, et comment les outils d'IA deviennent intégrés dans la pratique organisationnelle.

Les analyses quantitatives examinent comment les variations dans la représentation des données, la structure temporelle et l'architecture des modèles influencent la performance prédictive dans des contextes B2B. La construction des variables, les stratégies d'agrégation, les procédures d'échantillonnage et les choix algorithmiques sont systématiquement variés à travers des modèles d'apprentissage automatique et d'apprentissage profond. L'évaluation combine la validation croisée et des tests sur échantillon de validation indépendante (holdout), complétés par des ajustements fins (fine-tuning) et des analyses de seuils afin d'évaluer la pertinence des prédictions pour la prise de décision opérationnelle. Par cette approche, les volets quantitatifs dépassent le simple benchmarking de performance pour examiner comment les décisions fondées sur les données façonnent la pertinence pratique, la robustesse et l'interprétabilité managériale des modèles prédictifs.

L'analyse qualitative adopte une approche d'étude de cas imbriquée afin de comprendre comment les systèmes d'IA sont définis, développés et intégrés dans les processus organisationnels. La collecte de données s'appuie sur des entretiens, de la documentation interne et des traces des activités de développement. L'analyse progresse de manière itérative, en commençant par une reconstruction chronologique du projet, avant de passer à un codage systématique visant à identifier les thèmes récurrents, les mécanismes de coordination et les schémas décisionnels entre groupes de parties prenantes. La synthèse de ces schémas fournit des éclairages sur la manière dont le travail technique interagit avec les routines organisationnelles, comment la gouvernance et la collaboration influencent l'évolution de l'outil, et comment les processus d'intégration se déploient dans le temps.

Ensemble, ces composantes méthodologiques établissent une conception analytique cohérente et multiniveau. Les études quantitatives montrent comment les choix de modélisation et les structures de données façonnent le comportement prédictif et la pertinence pour la décision, tandis que l'étude qualitative clarifie la manière dont ces systèmes techniques interagissent avec les structures, capacités et pratiques organisationnelles lors de l'implémentation. Cette complémentarité permet à la thèse d'examiner l'IA simultanément comme artefact analytique et comme processus sociotechnique, en capturant l'interaction entre données, modèles et dynamiques organisationnelles qui sous-tend la création de valeur dans les environnements B2B à forte intensité relationnelle.

1.1.4 Cadre empirique

Cette thèse comprend trois études empiriques, chacune constituant un chapitre distinct, menées dans un environnement logiciel B2B. Ensemble, elles fournissent la base empirique du cadre analytique présenté ci-dessus en examinant comment l'IA crée de la valeur économique, rencontre des contraintes analytiques et organisationnelles, et devient intégrée dans la pratique organisationnelle.

Le Chapitre 2 s'appuie sur des données issues d'un fournisseur européen de logiciels B2B en mode Software-as-a-Service (SaaS), proposant un logiciel d'automatisation sous forme d'abonnements annuels renouvelables. Le jeu de données comprend des abonnements observés entre avril 2019 et juillet 2022. Chaque enregistrement d'abonnement inclut des informations firmographiques, l'historique transactionnel, les interactions avec le support client, ainsi que des données d'usage détaillées capturant les interactions des utilisateurs avec le logiciel. L'étude structure ces informations en variables prédictives selon trois

dimensions principales : temporalité, granularité et expertise. Ces variables structurées sont ensuite utilisées pour entraîner et évaluer une série de modèles d'apprentissage automatique pour la prédiction du churn client. L'analyse empirique intègre la préparation des données, l'ingénierie de variables et la validation des modèles à l'aide d'indicateurs de performance standard tels que l'aire sous la courbe ROC (AUC), le top-decile lift (TDL) et le profit maximum attendu (EMPB).

Le Chapitre 3 repose sur des données collectées auprès d'un fournisseur SaaS B2B. Le jeu de données inclut des demandes d'essai gratuit collectées entre avril 2022 et décembre 2024. Chaque essai donne un accès complet au logiciel pendant une période de 30 jours, après laquelle les prospects peuvent passer à un abonnement payant dans une fenêtre d'observation de 90 jours. Les données se composent de deux éléments principaux : d'une part, des informations sur les prospects obtenues via les formulaires d'inscription, incluant la région géographique, l'intention d'usage déclarée et d'autres caractéristiques ; d'autre part, des journaux d'usage détaillés et horodatés capturant l'ensemble des actions réalisées durant la période d'essai. Les données d'usage incluent des variables relatives aux niveaux d'activité, aux périodes de dormance, à la variété des actions, à la fréquence d'utilisation et à l'engagement des utilisateurs sur six catégories fonctionnelles du logiciel. Ces journaux sont utilisés pour construire des variables à la fois invariantes dans le temps et temporellement variables, permettant l'analyse des schémas d'usage quotidiens et des dynamiques de conversion. Le jeu de données permet également de simuler différentes stratégies d'intervention marketing.

Le Chapitre 4 repose sur une étude qualitative longitudinale d'un projet de développement d'IA mené entre 2019 et 2023 chez un fournisseur européen de logiciels B2B. L'étude se concentre sur le développement d'un système de rétention client basé sur l'apprentissage automatique, intégré à la plateforme de gestion de la relation client (CRM) de l'entreprise. La collecte de données combine plusieurs sources, dont 27 entretiens semi-directifs avec des experts techniques, des experts métiers et des partenaires de distribution ; des courriels internes et des comptes rendus de réunions documentant l'évolution du projet ; ainsi que des observations directes lors d'événements d'entreprise. L'analyse suit trois phases principales du projet : compréhension et préparation des données, définition du modèle et intégration de l'outil, retraçant la manière dont les processus techniques et organisationnels ont interagi tout au long du développement et du déploiement du système.

Collectivement, ces trois contextes empiriques couvrent à la fois des perspectives centrées sur les modèles et des perspectives organisationnelles, en englobant des aspects variés des outils d'IA utilisés dans des environnements logiciels B2B : analytique prédictive pour la rétention et la conversion, et évolution longitudinale d'un système d'IA au sein des structures internes et interorganisationnelles de l'entreprise.

1.1.5 Contributions

Création de valeur fondée sur les données dans les contextes B2B

La première contribution de cette thèse porte sur le développement et l'application de modèles d'IA qui créent de la valeur économique dans les environnements B2B. Les Chapitres 2 et 3 opérationnalisent cette dimension en démontrant comment des techniques analytiques avancées et de nouvelles sources de données élargissent le champ d'application de l'IA dans des contextes de prédiction business. Le Chapitre 4 complète cette perspective en examinant comment ces applications sont définies, adaptées et intégrées dans les structures organisationnelles au cours du développement de l'IA.

Le Chapitre 2 fait progresser la littérature sur la prédiction du churn client (CCP) et l'analytique marketing fondée sur les données en B2B, en clarifiant la manière dont l'apprentissage automatique peut transformer les données comportementales en informations actionnables. La recherche CCP en B2B s'est principalement concentrée sur le benchmarking algorithmique et s'est appuyée sur des variables

firmographiques, transactionnelles ou liées au support (Jahromi et al., 2014; Gordini and Veglio, 2017; Gattermann-Itschert and Thonemann, 2022; Janssens et al., 2022; Chen et al., 2015). Bien que ces travaux montrent que la modélisation prédictive améliore la gestion de la rétention, ils n'ont généralement pas intégré des informations d'usage capturant la manière dont les clients interagissent avec les produits et services numériques. Comme le soulignent Lilien (2016) et Wiersema (2013), bien que les entreprises B2B collectent de plus en plus de données comportementales volumineuses, ces données sont rarement structurées ou exploitées efficacement pour prédire le churn ou orienter les stratégies de gestion de la relation. Ce chapitre comble ce manque en proposant un cadre systématique pour intégrer les données d'usage dans les modèles de CCP, structuré selon trois dimensions analytiques : temporalité, granularité et expertise métier. Les tests empiriques portant sur des variables firmographiques, transactionnelles, de support et comportementales montrent que les données d'usage améliorent significativement à la fois la précision des modèles et leur interprétabilité, étendant les frontières de la création de valeur fondée sur les données dans l'analytique B2B. Sur le plan managérial, le cadre fournit aux entreprises SaaS une approche reproductible pour opérationnaliser les données d'usage produit dans la gestion de la rétention, permettant l'identification plus précoce des clients à risque, un engagement plus ciblé et une allocation plus efficace des ressources.

Le Chapitre 3 prolonge la recherche sur l'acquisition de clients et la prédiction de conversion de leads par l'IA dans les contextes SaaS B2B (Järvinen and Taiminen, 2016; Li, 2022; Yoganarasimhan et al., 2023). Les travaux existants s'appuient souvent sur des mesures d'usage d'essai statiques ou agrégées, telles que la fréquence totale d'activité ou la diversité, négligeant ainsi l'évolution dynamique de l'usage durant les périodes d'évaluation d'essai (Li, 2022; Yoganarasimhan et al., 2023). Cette limitation restreint la compréhension de la manière dont les clients potentiels apprennent, révisent leurs croyances et signalent leur intention d'achat pendant les essais gratuits. Le Chapitre 3 répond à ce manque en développant un cadre prédictif dynamique qui intègre des variables d'usage temporellement variables afin de capturer les séquences d'engagement pendant l'essai. En comparant des architectures d'apprentissage automatique traditionnelles et d'apprentissage profond dans une évaluation à horizon glissant, l'étude montre comment l'évolution des données d'usage améliore la prédiction précoce de conversion et l'optimisation des ressources marketing. Cette approche intègre la temporalité et l'apprentissage dans la modélisation prédictive, montrant que la valeur issue de l'IA provient également de la modélisation de l'évolution de l'usage, et pas seulement de l'obtention d'une précision élevée à un instant final. Dans les applications SaaS B2B, l'intégration de ces dynamiques temporelles soutient une qualification plus précoce des leads, l'optimisation du moment d'outreach et des interventions d'onboarding visant à maximiser l'efficacité de conversion.

Enfin, le Chapitre 4 contribue à la littérature qui relie la conception de systèmes modulaires en systèmes d'information et en théorie des organisations à l'usage croissant des applications d'IA fondées sur les données dans les entreprises (Campagnolo and Camuffo, 2010; Baldwin and Clark, 2000; Ravasi and Stigliani, 2012). Les travaux antérieurs montrent que les technologies complexes sont le plus souvent développées lorsque les architectures techniques et les divisions organisationnelles du travail sont alignées, une relation décrite dans la littérature comme la correspondance entre « ce qui dépend de quoi » et « qui fait quoi » (Colfer and Baldwin, 2016; Baldwin, 2023; Puranam et al., 2012). Bien que les avancées en collecte de données et en capacité de calcul aient permis aux entreprises de concevoir et d'intégrer des modèles prédictifs d'IA (Ma et al., 2024; Chen et al., 2024; Wang et al., 2022), la recherche existante a accordé une attention limitée à la manière dont ces applications sont définies et réalisées via des processus coordonnés de données et d'organisation. Cette étude comble ce manque en montrant que, dans les contextes B2B, les applications d'IA émergent par étapes de développement modulaires, au cours desquelles les exigences techniques pour l'intégration des données génèrent des ajustements organisationnels correspondants (Shrestha et al., 2021; Sjödin et al., 2021). Ainsi, le potentiel de création de valeur de l'IA dépend de la manière dont ses composants techniques modulaires s'alignent sur les structures et

les flux organisationnels, et les reconfigurent. Pour les managers, cela souligne que la définition des applications d'IA implique non seulement l'identification des besoins business, mais aussi la conception de la gouvernance des données, l'allocation des rôles et la collaboration interfonctionnelle afin de tirer pleinement parti de la modularité dans les environnements B2B.

Les trois chapitres saisissent collectivement la manière dont l'IA crée, contraint et soutient la valeur à différents niveaux du cycle de vie relationnel B2B. Le Chapitre 3 traite de l'amont de la création de valeur en examinant comment l'analytique prédictive soutient l'acquisition de clients par la modélisation dynamique de la conversion. Le Chapitre 2 se concentre sur l'aval, en analysant comment l'apprentissage automatique améliore la rétention client en identifiant et en prévenant le churn. Le Chapitre 4 étend l'analyse au niveau de l'intégration de l'IA, en étudiant comment les organisations développent et gouvernent les capacités qui permettent à ces applications de fonctionner de manière fiable dans le temps. Plutôt que de correspondre à des étapes marketing distinctes, ces chapitres représentent des mécanismes interdépendants au sein d'un écosystème piloté par l'IA en évolution, reliant les applications orientées client aux infrastructures organisationnelles qui les soutiennent.

Contraintes analytiques, opérationnelles et organisationnelles

La deuxième contribution de cette thèse réside dans l'identification et l'analyse des défis qui limitent l'intégration de l'IA dans les contextes B2B. À travers les trois études, ces défis émergent à plusieurs niveaux — technique, analytique, managérial et organisationnel — révélant les conditions dans lesquelles le potentiel analytique de l'IA rencontre des limites dans les environnements réels.

Dans le champ de l'apprentissage automatique et de l'analytique B2B, le Chapitre 2 fait également progresser la compréhension des contraintes analytiques et opérationnelles qui façonnent l'intégration de l'IA dans les environnements business. Les travaux antérieurs ont reconnu la difficulté technique d'intégrer des données comportementales hétérogènes et de haute dimension dans les architectures d'apprentissage, mais offrent peu de recommandations sur la manière de structurer ces données pour atténuer la rareté, l'incohérence et la complexité de prétraitement (Coussement and Van den Poel, 2008). Cette étude donne un cadre pour structurer les données d'usage selon trois dimensions analytiques, transformant ainsi une source de données peu explorée et techniquement exigeante en une ressource pouvant être efficacement opérationnalisée pour l'apprentissage automatique. Outre la résolution de problèmes de tractabilité analytique, ce chapitre introduit une perspective opérationnelle complémentaire en quantifiant l'empreinte carbone et les coûts computationnels associés à l'entraînement et à l'évaluation des modèles. Cette analyse met en lumière la durabilité comme une contrainte critique, mais souvent négligée, des expérimentations d'IA à grande échelle. Collectivement, ces contributions redéfinissent l'intégration de l'IA comme un processus devant équilibrer rigueur analytique, efficacité des ressources et responsabilité environnementale.

Le Chapitre 3 contribue également à la littérature sur la méthodologie de l'IA et de l'analytique prédictive, en particulier sur le compromis entre sophistication des modèles et interprétabilité managériale dans la prise de décision B2B (Meire et al., 2017; González-Flores et al., 2025). Les travaux antérieurs ont mis en avant les gains prédictifs des modèles complexes, mais leurs sorties ne sont pas systématiquement comprises dans la pratique organisationnelle, surtout lorsque les décisions dépendent de la confiance et de la compréhension des utilisateurs non techniques. Ce déficit d'interprétabilité constitue un manque théorique et opérationnel, limitant l'intégration de l'IA malgré des performances techniques élevées. Le Chapitre 3 comble ce manque en transformant les sorties des modèles en trajectoires d'engagement interprétables qui capturent les schémas d'usage des prospects, reliant les prédictions à des profils de prospects reconnaissables. Cette contribution fait progresser la littérature en démontrant une voie permettant de maintenir la précision tout en renforçant la compréhensibilité, tant sur le plan analytique

que managérial. De plus, l'étude fournit aux managers des profils d'usage empiriquement dérivés qui facilitent la priorisation fondée sur les données dans les campagnes de ciblage, permettant aux équipes commerciales et marketing d'opérationnaliser les insights générés par l'IA avec davantage de confiance et de redevabilité.

Le Chapitre 4 contribue à la littérature sur les défis et la gouvernance de l'IA dans les organisations, qui examine les barrières empêchant une intégration scalable et fiable de l'IA (Mikalef and Gupta, 2021; Weber et al., 2023; Schneider et al., 2023). Les travaux antérieurs mettent en évidence des problèmes tels que la qualité limitée des données, l'opacité méthodologique et l'insuffisance de la collaboration interfonctionnelle, mais tendent à analyser ces contraintes de manière isolée, sans considérer comment elles coévoluent et interagissent durant le processus de développement de l'IA. Cette étude comble ce manque en montrant que les contraintes analytiques, opérationnelles et organisationnelles sont interdépendantes et cumulatives, émergeant séquentiellement au fil des phases de compréhension et préparation des données, de définition du modèle et d'intégration de l'outil. En mobilisant le concept d'hypothèse de mirroring, la recherche montre que le développement de l'IA reproduit les structures organisationnelles et schémas de communication existants, faisant des ruptures de coordination et des inadéquations de gouvernance des obstacles centraux à l'intégration. Ce faisant, l'étude redéfinit l'intégration de l'IA comme un défi processuel et systémique plutôt qu'un ensemble de problèmes techniques isolés. Pour les praticiens, ces résultats soulignent que surmonter les contraintes liées à l'IA exige une coordination interfonctionnelle continue, des mécanismes de gouvernance adaptatifs et une flexibilité structurelle permettant de gérer les interdépendances entre conception technique et pratique organisationnelle.

Insertion et pérennisation de l'IA dans la pratique organisationnelle

La troisième contribution porte sur la manière dont les systèmes d'IA sont implémentés et pérennisés dans la pratique organisationnelle. Cette contribution va au-delà du développement de modèles ou de l'identification de barrières pour examiner les mécanismes par lesquels l'IA devient opérationnelle, routinisée et continûment génératrice de valeur au sein des entreprises. Le Chapitre 4 fait progresser la littérature sur la conception organisationnelle et la co-création de l'IA, qui étudie la coévolution des systèmes technologiques et organisationnels durant l'implémentation de l'IA (Bailey and Barley, 2020; Lebovitz et al., 2021; Van den Broek et al., 2021; Waardenburg and Huysman, 2022). La recherche existante a principalement traité le mirroring entre technologie et organisation comme une relation statique (Colfer and Baldwin, 2016; Burton and Galvin, 2022), offrant une compréhension limitée de la manière dont cette relation se déploie dans le temps ou avec de multiples parties prenantes. En répondant à ce manque théorique, l'étude propose un récit processuel du mirroring, montrant comment les exigences technologiques et organisationnelles coévoluent au fil de trois phases. Les résultats révèlent que le processus de mirroring est itératif, les hypothèses organisationnelles et la participation limitée des utilisateurs finaux tendant à s'inscrire dans l'outil final, ce qui réduit son utilisabilité et son alignement avec les pratiques de travail. Cette perspective redéfinit l'ancrage de l'IA comme un processus de co-création façonné par la co-conception, l'apprentissage et l'adaptation. D'un point de vue managérial, pérenniser l'IA dans la pratique nécessite de cultiver des capacités de coordination dynamiques, une implication précoce et équilibrée des parties prenantes, ainsi que des boucles de rétroaction continues alignant les fonctionnalités de l'outil avec l'évolution des besoins des utilisateurs et des objectifs organisationnels.

Le Chapitre 3 contribue en outre à la littérature sur l'implémentation de l'IA et le marketing fondé sur les données, en démontrant comment l'analytique prédictive peut être opérationnalisée pour guider la prise de décision en temps quasi réel et l'allocation des ressources dans les contextes B2B (Blattberg et al., 2008; Li et al., 2016). Alors que les travaux existants sur la gestion des leads pilotée par l'IA ont principalement mis l'accent sur la précision prédictive, cette étude fait progresser la compréhension de la

manière dont les sorties prédictives se traduisent en impact économique via une analyse décisionnelle basée sur la simulation. En intégrant des prévisions hebdomadaires de modèles à un cadre de simulation de profit, l'étude quantifie les retours financiers attendus de différentes stratégies d'outreach, montrant que les interventions guidées par les modèles augmentent substantiellement les revenus liés à la conversion tout en réduisant l'engagement inefficace auprès de prospects. Cette approche positionne l'implémentation de l'IA comme un processus de calibration continue de la valeur, dans lequel prédiction, action et rétroaction sont systématiquement reliées. L'étude comble le fossé entre la prédiction algorithmique et le pilotage managérial en intégrant l'évaluation économique au cœur du flux de travail de l'IA. Pour les praticiens, les résultats soulignent que la réalisation du potentiel financier de l'IA exige l'alignement des processus analytiques avec le rythme opérationnel, en veillant à ce que les actions guidées par les modèles soient synchronisées avec la planification commerciale, le calendrier des campagnes et les contraintes de ressources, afin de maximiser le retour sur investissement marketing.

1.2 General Introduction

1.2.1 Context and Motivation

The contemporary business environment is defined by an unprecedented proliferation of digital data sources, creating complex information ecosystems in which organizations face continuous streams of operational, transactional, and interactional data (Mcafee and Brynjolfsson, 2012; Grover et al., 2018; Kellogg et al., 2020). The resulting growth in data volume, velocity, and variety has shifted the strategic challenge from data acquisition to the extraction of actionable intelligence from high-dimensional, heterogeneous datasets that exceed human cognitive and traditional analytical capacities (Kellogg et al., 2020; George et al., 2014).

This complexity exposes the limitations of conventional analytical approaches. Traditional statistical methods and rule-based decision support systems, designed for structured data and stable patterns, cannot effectively process current information streams. Recognizing this gap, organizations have increasingly turned to artificial intelligence (AI) as a transformative solution. Recent advances in machine learning (ML), deep learning (DL), and natural language processing (NLP) have demonstrated remarkable capabilities, from surpassing human performance in image recognition to generating sophisticated language models (Jordan and Mitchell, 2015; Lecun et al., 2015; Krizhevsky et al., 2017). These technical breakthroughs have catalyzed extensive AI integration across industries, with digital-native companies and consumer-facing applications leading the transformation (Mcafee and Brynjolfsson, 2012).

AI has already generated substantial value across diverse sectors. Platform-based firms have captured billions in revenue through recommendation systems and targeted advertising (Gomez-Uribe and Hunt, 2015; Covington et al., 2016; Linden et al., 2017). Financial services organizations have automated complex processes with measurable returns (Bahoo et al., 2024), while retailers employ predictive analytics to optimize operations and customer experiences (Fildes et al., 2022). These cases illustrate AI's transformative potential for business, yet the extent of realized benefits depends on context-specific conditions and purpose-specific objectives (Ledro et al., 2025).

Business-to-business (B2B) contexts represent a setting in which these conditions become particularly pronounced. In B2B markets, relationships are not discrete transactions but long-term, trust-based exchanges that require ongoing commitment and collaboration (Morgan and Hunt, 1994). B2B firms often rely on complex decision-making networks and professional ties that extend beyond formal contractual boundaries (Mojir and Anbil, 2025).

The relational nature of B2B settings has direct implications on how AI can be integrated into these settings. Extended sales cycles reduce the frequency of observable outcomes and thereby limit the scope for empirical validation of AI-generated insights (Moradi and Dass, 2022; Syam and Sharma, 2018). Decision-making processes involving multiple stakeholders introduce ambiguity in value attribution, as predictive outputs are interpreted and utilized by diverse actors across industry levels (Purmonen et al., 2023). Moreover, AI-driven decision support depends on the availability, structuring, and integration of often sparse and relationship-dependent data, imposing additional demands on model specification and evaluation in B2B environments (Lim et al., 2019).

Taken together, these conditions position B2B settings as a critical context for examining AI integration, where predictive systems must operate within networks of trust, sustained interaction, and strategic interdependence. Within these relationship-dependent contexts, the integration of AI further raises fundamental questions concerning data availability and feedback delays, value attribution across multiple decision units, the interoperability of new systems with legacy infrastructures, and the governance of human–AI role allocation (Nyadzayo et al., 2020). Together, these relational and technical

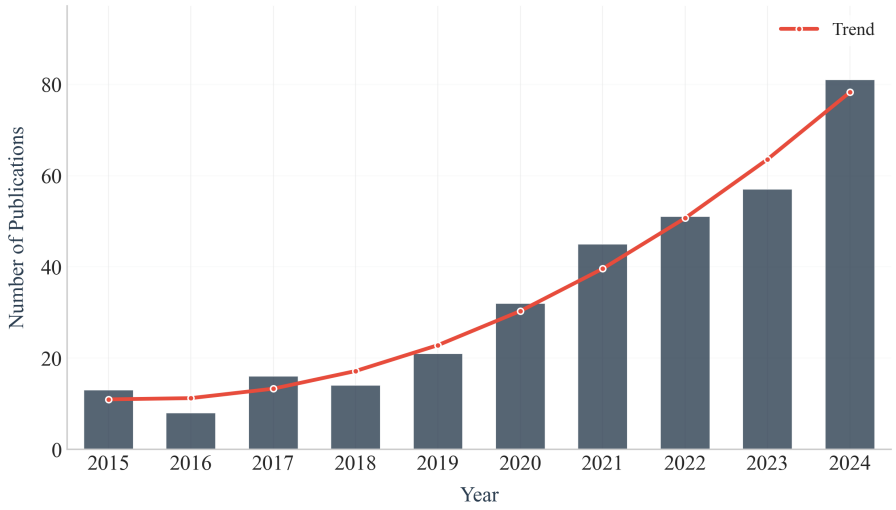


Figure 1.4: Temporal distribution of B2B AI publications

complexities underscore the importance of developing analytical approaches and predictive systems that can accommodate the dynamics of B2B interactions, alongside the organizational and data constraints, while supporting responsible and relationship-preserving AI adoption.

To situate this work within the broader scholarly landscape, a structured review of peer-reviewed publications on AI in B2B contexts between 2015 and 2024 was conducted. The literature reveals a clear and accelerating growth trajectory, with contributions increasing substantially after 2019 and AI in B2B transitioning from a marginal topic to a recognized research area. As shown in Figure 1.4, publication activity has risen steadily since 2019, reflecting the emergence of B2B AI as an established academic domain. Over time, the thematic emphasis has also shifted from early function-specific applications, such as sales forecasting and demand prediction, toward more technology-centered perspectives and, more recently, toward diversified analytical methods and strategic domains. This evolution, summarized in Figure 1.5, depicts the transition from operationally oriented studies to research engaging with advanced AI techniques and broader business questions.

Despite this growth, the existing literature presents several limitations. Many contributions illustrate use cases in domains such as sales forecasting, customer management, and marketing analytics (Gordini and Veglio, 2017; Jahromi et al., 2014; Gattermann-Itschert and Thonemann, 2022). Recent work has extended this view by mapping AI to the customer lifecycle and by documenting industry perceptions of benefits, barriers, and opportunities (Moradi and Dass, 2022). These studies establish AI's potential in B2B and demonstrate clear pathways to enhanced efficiency and customer engagement. Yet they remain concentrated in a narrow set of functional domains and often rely on similar data sources, overlooking how novel types of data combined with advanced techniques such as deep learning, can reshape value creation in B2B settings. The gap therefore lies in extending the evidence base toward practical applications that exploit novel data sources and explicitly link AI models to measurable business outcomes, thereby broadening our understanding of AI's contribution to B2B performance.

A second limitation concerns the fragmentation of research on constraints to AI integration. While prior research has identified technical and organizational constraints to AI integration, it lacks an integrated perspective. Technical challenges such as data quality, access, and model robustness (Dwivedi and Wang, 2022) often appear disconnected from organizational barriers, including capability shortages, cultural resistance, and resource constraints. In B2B marketing in particular, the high costs of advanced AI systems and the difficulty of aligning them with heterogeneous datasets and multi-stakeholder decision structures pose additional obstacles (Chatterjee et al., 2024). However, these barriers are frequently

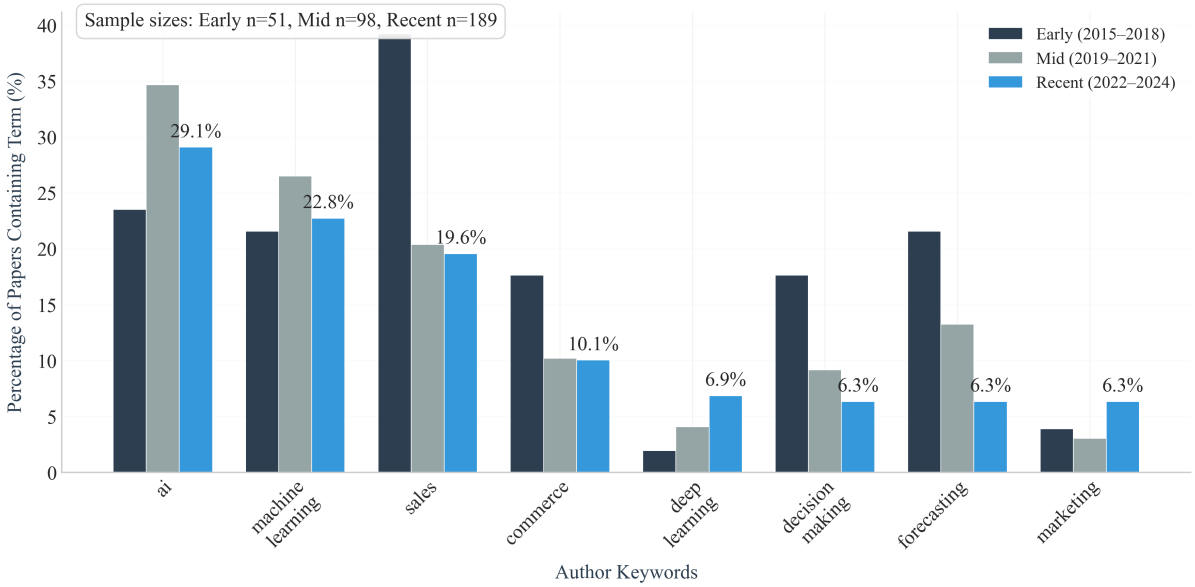


Figure 1.5: Temporal Evolution of Author Keywords Across Research Periods

considered in isolation, without linking them to the technical configurations of AI systems or to the business value they are designed to produce.

A third limitation relates to the organizational processes shaping AI implementation. Much of the existing literature emphasizes proof-of-concept demonstrations or pilot applications, providing limited insight into how AI systems are adapted, governed, and institutionalized within firms. While prior work highlights AI's potential to transform sales, marketing, customer engagement, and inter-firm collaboration (Paschen et al., 2019), and recent studies acknowledge challenges of alignment and integration (Keegan et al., 2024), systematic analyses of how AI interacts with extended sales cycles and multi-stakeholder decision processes are still scarce. Developing such understanding is essential for explaining how AI evolves from initial experimentation to sustained use in B2B contexts.

Taken together, these gaps highlight the need for a more integrated understanding of how AI can create value, how methodological choices shape predictive performance, and how organizational conditions influence the successful adoption of AI in B2B contexts. The purpose of this thesis is to address these limitations by examining the design, evaluation, and organizational integration of predictive systems in relationship-intensive B2B environments. To advance this understanding, the dissertation is guided by three research questions. First, how can different AI methods and data configurations be designed and evaluated to generate predictive and operational value in B2B settings? Second, what technical and organizational conditions constrain the scalability, reliability, and value realization of AI applications in B2B contexts? Third, how do AI applications become embedded, adapted, and institutionalized within B2B organizations over time, and how do these implementation processes interact with emerging challenges and analytical configurations? The chapters that follow address these questions through a combination of quantitative and qualitative analyses, providing a multi-level understanding of how AI evolves from initial application to sustained integration in B2B environments.

1.2.2 Analytical Framework

Building on these research questions, the dissertation introduces an analytical framework that conceptualizes AI integration in B2B settings as a continuous process unfolding across three interconnected dimensions: Applications, Challenges, and Implementation. This framework structures the analysis

required to examine how applications of predictive systems are designed and evaluated (RQ1), how technical and organizational challenges shape their scalability and value realization (RQ2), and how AI tools become embedded, adapted, and institutionalized within firms (RQ3). Organizing the thesis around these dimensions provides a coherent basis for analyzing the technological, analytical, and organizational processes through which AI evolves and creates value in relationship-intensive B2B environments.

The framework draws on prior empirical and theoretical work to capture the interaction between technological mechanisms and organizational dynamics. Each dimension offers a distinct analytical lens, and together they support the dissertation's layered research design, where quantitative analyses evaluate model performance and the implications of different data configurations, and qualitative analyses contextualize and interpret them within organizational practice.

The *applications* dimension examines how AI techniques generate business value when aligned with B2B tasks and data. Prior research suggests that AI value impact depends on the fit between method, data structure, and decision context. For instance, Syam and Sharma (2018) show how AI augments B2B sales management through dynamic lead scoring and opportunity qualification by embedding predictions into seller workflows. Similarly, Chatterjee et al. (2022) find that AI-enabled CRM systems in industrial markets are associated with greater relationship satisfaction and improved firm performance, underscoring the importance of integrating analytical outputs with customer-facing routines. Extending these domain-specific studies, Vlačić et al. (2021) conduct a systematic review of AI applications in marketing, demonstrating that machine learning and deep learning expand analytical capability by processing large, heterogeneous datasets and uncovering patterns in customer and market behavior that exceed the limits of traditional statistical models. The study argues that these techniques create value only when aligned with data availability, organizational readiness, and managerial interpretability, conditions that determine whether advanced analytical methods become actionable within firms.

Within this framework, *applications* represent the technological mechanisms through which AI transforms organizational operations and creates business value (Enholm et al., 2022). It provides an analytical lens for examining how different AI methodologies interact with the data environments characteristic of B2B organizations. Quantitative analyses contribute by assessing how model architectures, training processes, and emerging data sources reshape prediction, and decision-making support, while qualitative approaches reveal how these outputs are defined, appropriated, integrated, and institutionalized within organizational practice.

The second dimension, *challenges*, examines the conditions that constrain AI integration and limit the consistent realization of business value. These challenges arise from both technical and organizational sources that shape how effectively analytical capabilities can be translated into operational outcomes. From a methodological perspective, research underscores how the proliferation of unstructured and heterogeneous data has fundamentally reshaped the analytical landscape, prompting the development of advanced algorithms and knowledge representation techniques to manage this complexity (Agarwal and Dhar, 2014). Yet, as Mikalef and Gupta (2021) emphasize, in B2B contexts the limited availability of large-scale, high-granularity data constrains the development of robust and generalizable models, thereby limiting the extent to which firms can translate AI's analytical potential into sustained value creation.

At the organizational level, studies emphasize governance, coordination, and capability limitations that impede the scalability of AI initiatives. In this context, Weber et al. (2023) identify effective cross-functional collaboration as a critical organizational challenge, noting that the inherent inscrutability of AI systems amplifies the information gap between technical departments and other business functions. Similarly, Schneider et al. (2023) emphasize that organizations encounter persistent difficulties in establishing robust data governance, developing sufficient analytical capabilities, and ensuring coordination across functional units, all of which constrain the effective scaling and realization of AI-driven value. In B2B marketing, Keegan et al. (2024) illustrate that AI integration is often hindered by fundamental contradictions between

solution providers and client firms. These contradictions manifest through misaligned value systems and a lack of shared understanding of how AI can create value, disruptions to established work practices as AI technologies are assimilated, and discrepancies between the expected and realized benefits of AI among managerial stakeholders.

Within this framework, the *challenges* dimension offers a conceptual lens for examining the conditions under which AI integration in B2B contexts fails to achieve consistency or scalability. Research increasingly shows that the progression from analytical potential to realized business value is nonlinear, shaped by the mutual dependence of technical and organizational systems. The *challenges* dimension captures these points of friction, where misalignments between methodological rigor, data infrastructures, and organizational capabilities disrupt the scalability and reliability of AI outcomes. Quantitative approaches reveal the methodological vulnerabilities that constrain model robustness and generalizability, while qualitative analyses uncover governance, coordination, and interpretive tensions that hinder the integration of AI within organizational practice. Together, these perspectives advance an integrated understanding of AI as a socio-technical system, in which challenges constitute inherent features that define the boundaries of sustainable value creation in B2B environments.

The third dimension, *implementations*, concerns the processes through which AI applications are adopted and institutionalized within organizations. As firms integrate AI into decision processes, new patterns of collaboration, responsibility, and control emerge (Von Krogh, 2018). Research further indicates that effective implementation depends on developing complementary organizational capabilities that facilitate integration across firm boundaries. In B2B contexts, Dubey et al. (2021) emphasize the importance of alliance management capability, which enables organizations to coordinate partnerships and share strategic resources essential for realizing AI's full potential.

Within this framework, the *implementations* dimension captures the evolving interplay between technological systems and organizational structures as AI becomes embedded in business processes. It focuses on how implementation unfolds through cycles of adaptation, learning, and coordination that redefine roles, workflows, and governance arrangements. A technical perspective enables the assessment of how AI influences performance and operational efficiency, while an organizational perspective captures the evolution of routines, decision practices, and governance mechanisms that form around new technological capabilities. Together, these perspectives illuminate the processes through which AI extends beyond its technical function to become an embedded organizational force in B2B settings.

Although analytically distinct, the three dimensions of Applications, Challenges, and Implementation are empirically intertwined and mutually reinforcing. Prior studies have typically examined these domains in isolation—addressing algorithmic performance, organizational barriers, or adoption processes separately—but increasing evidence indicates that they co-evolve as part of a recursive process of AI integration (George et al., 2014; Mariani et al., 2023). In B2B environments, where data are distributed across organizational boundaries and decision-making involves multiple actors, this recursive character is particularly salient. The deployment of new AI applications often gives rise to unanticipated challenges, such as heightened dependencies on client-generated data, the need for interpretability in complex sales or service processes, and emerging governance tensions over data ownership and accountability (Moradi and Dass, 2022). Efforts to address these issues through improved data infrastructures, enhanced model transparency, and the development of cross-functional analytical capabilities, create the conditions that enable further application and refinement, reinforcing an ongoing cycle of technological and organizational adaptation.

As AI systems are integrated, they reshape the informational and organizational environments that initially constrained them. In B2B contexts, these feedback effects are amplified by extended sales cycles, service dependencies, and inter-firm coordination mechanisms. This recursive pattern is also reflected in the methodological foundations of AI research itself. Studies of machine learning lifecycles emphasize

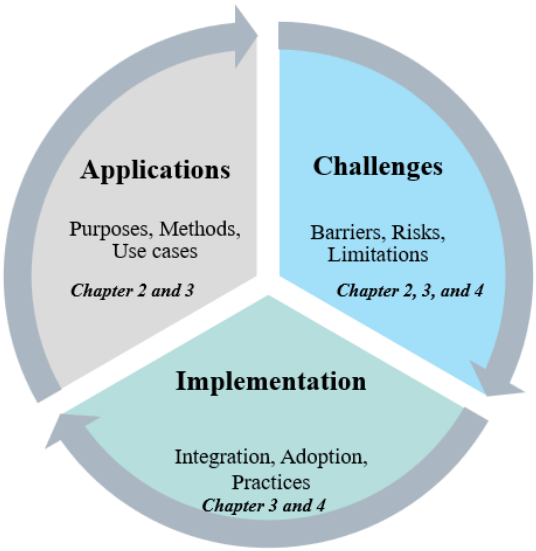


Figure 1.6: Conceptual Framework: Interactions between Applications, Challenges, and Implementation

that model reliability and value realization depend on iterative refinement and contextual feedback rather than static design (Sculley et al., 2015). Maintaining performance requires continual adjustment to heterogeneous data inputs, evolving customer portfolios, and shifting business objectives. Research on AI reinforces this alignment by emphasizing that progress increasingly depends on improving the quality, representativeness, and governance of data rather than on more complex model architectures.

As depicted in Figure 1.6, the three dimensions are not treated as separate categories but as interdependent elements of a continuous process. The framework conceptualizes AI integration in B2B contexts as a cumulative and evolving process. It provides a foundation for analyzing how applications, technical and organizational constraints, and implementation processes develop and shape one another over time. Methodologically, the layered research design reinforces this recursive view by connecting quantitative and qualitative approaches within a unified analytical logic. Together, these layers enable a multi-level understanding of AI integration, tracing how technological capability, analytical complexity, and organizational practice co-evolve in the creation and stabilization of business value within B2B settings.

In sum, by framing the three dimensions as interconnected rather than sequential, the framework emphasizes that effective AI integration in B2B settings depends not on any single factor in isolation but on the synergistic relationships among them. It invites both researchers and practitioners to move beyond compartmentalized analyses to a integrated perspective on how AI systems evolve as part of the organizational and inter-organizational fabric. In doing so, it advances a richer and more actionable understanding of the socio-technical and methodological dynamics through which AI's promise is realized, adapted, and sustained in complex B2B environments.

To demonstrate how the conceptual framework structures the empirical analysis, the three dimensions are operationalized across the dissertation in an overlapping yet structured manner. The Applications dimension is examined in Chapters 2 and 3, which investigate how different AI methods and data configurations generate value in B2B retention and acquisition contexts. The Challenges dimension spans Chapters 2 through 4, as analytical, operational, and organizational constraints emerge across the development, evaluation, and deployment of AI systems. The Implementation dimension is addressed primarily in Chapter 4, with additional practical insights from Chapter 3, analyzing how AI applications become adopted, adapted, and sustained within organizational processes. Together, these chapters

collectively operationalize the framework, ensuring coherence between the conceptual model and the dissertation's empirical contributions.

1.2.3 Methodological approach

The dissertation follows a multi-method research approach that integrates quantitative machine-learning experimentation with qualitative analysis. This design enables the study of AI systems from complementary perspectives, how predictive value is generated through data and models, how constraints emerge during integration, and how AI tools become embedded in organizational practice.

The quantitative analyses investigate how variations in data representation, temporal structure, and model architecture influence predictive performance in B2B contexts. Feature construction, aggregation strategies, sampling procedures, and algorithmic choices are systematically varied across machine learning and deep learning models. The evaluation combines cross-validation with holdout tests, supported by fine-tunning and threshold analyses to assess the relevance of predictions for operational decision-making. Through this approach, the quantitative components extend from performance benchmarking to examining how data-driven decisions shape the practical relevance, robustness, and managerial interpretability of predictive models.

The qualitative analysis adopts an embedded case-study approach to understand how AI systems are defined, developed, and integrated within organizational processes. Data collection draws on interviews, internal documentation, and records of development activities. The analysis proceeds iteratively, beginning with a chronological reconstruction of the project before moving to systematic coding to identify recurring themes, coordination mechanisms, and decision patterns across stakeholder groups. The synthesis of these patterns provides insights into how technical work interacts with organizational routines, how governance and collaboration influence tool evolution, and how integration processes unfold over time.

Together, these methodological components establish a coherent, multi-level analytical design. The quantitative studies illuminate how modelling choices and data structures shape predictive behaviour and decision relevance, while the qualitative study clarifies how these technical systems interact with organizational structures, capabilities, and practices during implementation. This complementarity enables the dissertation to examine AI simultaneously as an analytical artefact and as a socio-technical process, capturing the interplay between data, models, and organizational dynamics that underlies value creation in relationship-intensive B2B environments.

1.2.4 Empirical setting

This dissertation comprises three empirical studies, each constituting a separate chapter, conducted in a B2B software environment. Together, they provide the empirical foundation for the analytical framework introduced above by examining how AI creates business value, encounters analytical and organizational constraints, and becomes embedded in organizational practice.

Chapter 2 draws on data from a European B2B Software-as-a-service (SaaS) provider offering automation software under renewable annual subscriptions. The dataset comprises subscriptions observed between April 2019 and July 2022. Each subscription record includes firmographic information, transactional history, customer support interactions, and detailed usage data capturing user interactions with the software. The study structures this information into predictive features across three main dimensions: timing, granularity, and expertise. These structured features are then used to train and evaluate a series of machine-learning models for customer churn prediction. The empirical analysis incorporates data preprocessing, feature engineering, and model validation using standard performance metrics such as the area under the ROC curve (AUC), top-decile lift (TDL), and expected maximum profit (EMPB).

Chapter 3 is based on data collected from a B2B SaaS provider. The dataset includes free-trial requests collected between April 2022 and December 2024. Each trial grants full access to the software for a 30-day period, after which prospects can upgrade to a paid subscription within a 90-day observation window. The data consist of two main components: prospect information obtained from registration forms, including geographic region, stated usage intention, and detailed time-stamped usage logs capturing all actions performed during the trial period. The usage data include variables related to activity levels, dormancy, action variety, frequency of use, and user engagement across six functional categories of the software. These logs are used to construct both time-invariant and time-varying variables, allowing the analysis of daily usage patterns and conversion dynamics. The dataset also supports the simulation of different marketing intervention strategies.

Chapter 4 is based on a qualitative, longitudinal study of an AI development project carried out between 2019 and 2023 at a European B2B software provider. The study focuses on the development of a machine-learning–based customer retention system integrated into the company's customer relationship management (CRM) platform. Data collection combines multiple sources, including 27 semi-structured interviews with technical experts, business experts, and channel partners; internal emails and meeting minutes documenting the project's evolution; and direct observations during company events. The analysis follows three main phases of the project, data understanding and preparation, model definition, and tool integration, tracing how technical and organizational processes interacted throughout the system's development and deployment.

Collectively, the three empirical settings encompass both model-based and organizational perspectives, covering diverse aspects of AI tools used in B2B software contexts: predictive analytics for retention and conversion, and the longitudinal evolution of an AI system within a firm's internal and interfirm structures.

1.2.5 Contributions

Data-Driven Value Creation in B2B Contexts

The first contribution of this dissertation concerns the development and application of AI models that create business value in B2B settings. Both Chapter 2 and Chapter 3 operationalize this dimension by demonstrating how advanced analytical techniques and novel data sources expand the scope of AI applications in predictive business contexts. Chapter 4 complements this perspective by examining how such applications are defined, adapted, and embedded within organizational structures during AI development.

Chapter 2 advances the B2B customer churn prediction (CCP) and data-driven marketing analytics literature by clarifying how machine learning can transform behavioral data into actionable insights. Prior CCP research in B2B settings has largely emphasized algorithmic benchmarking and relied on firmographic, transactional, or support-related variables (Jahromi et al., 2014; Gordini and Veglio, 2017; Gattermann-Itschert and Thonemann, 2022; Janssens et al., 2022; Chen et al., 2015). Although these studies show that predictive modeling improves retention management, they have failed to incorporate usage-level information that captures how customers interact with digital products and services. As Lilien (2016) and Wiersema (2013) note, while B2B firms increasingly collect vast behavioral data, such data are rarely structured or exploited effectively to predict churn or inform relationship management strategies. This chapter addresses that gap through a systematic framework for integrating usage data into CCP models, structured along three analytical dimensions, timing, granularity, and domain expertise. Empirical testing across firmographic, transactional, support, and behavioral variables shows that usage data significantly enhance both model accuracy and interpretability, extending the boundaries of data-

driven value creation in B2B analytics. On the managerial side, the framework provides SaaS firms with a replicable approach to operationalize product usage data for retention management, enabling earlier identification of at-risk customers, targeted engagement, and more efficient resource allocation.

Chapter 3 extends research on AI-enabled customer acquisition and lead conversion prediction in B2B SaaS contexts (Järvinen and Taiminen, 2016; Li, 2022; Yoganarasimhan et al., 2023). Existing studies often rely on static or aggregated trial-usage metrics, such as total frequency of activity or diversity, thus overlooking the dynamic evolution of usage during trial-evaluation periods (Li, 2022; Yoganarasimhan et al., 2023). This limitation constrains understanding of how prospective customers learn, update their beliefs, and signal purchase-intent during free trials. Chapter 3 addresses this gap by developing a dynamic predictive framework that integrates time-varying usage variables to capture trial engagement sequences. By comparing traditional machine-learning and deep-learning architectures within a rolling-horizon evaluation, the study demonstrates how evolving usage data improve early conversion prediction and marketing resource optimization. This approach embeds temporality and learning into predictive modeling, showing that AI-derived value also arises from modeling the evolution of usage in addition to achieving high endpoint accuracy. In B2B SaaS applications, incorporating these temporal dynamics supports earlier lead qualification, optimization of outreach time, and onboarding interventions to maximize conversion efficiency.

Finally, Chapter 4 contributes to the literature that connects modular system design in information systems and organizational studies with the growing use of data-driven AI applications in firms (Campagnolo and Camuffo, 2010; Baldwin and Clark, 2000; Ravasi and Stigliani, 2012). Prior research demonstrates that complex technologies are most frequently developed when technical architectures and organizational divisions of labour are aligned, a relationship described in literature as the correspondence between "what depends on what" and "who does what" (Colfer and Baldwin, 2016; Baldwin, 2023; Puranam et al., 2012). Although advances in data collection and computational power have enabled firms to design and integrate predictive AI models (Ma et al., 2024; Chen et al., 2024; Wang et al., 2022), existing research has paid limited attention to how these applications are defined and realized through coordinated data and organizational processes. This study addresses this gap by showing that, in B2B contexts, AI applications emerge through modular development stages where technical requirements for data integration generate corresponding organizational adjustments (Shrestha et al., 2021; Sjödin et al., 2021). Hence, AI's value-creation potential depends on how effectively its modular technical components align with, and reconfigure, organizational structures and workflows. For managers, this highlights that defining AI applications involves business needs identification but also the design of data governance, role allocation, and cross-functional collaboration to leverage modularity effectively within B2B environments.

The three chapters collectively capture how AI creates, constrains, and sustains value across different layers of the B2B relationship lifecycle. Chapter 3 addresses the front-end of value creation by examining how predictive analytics supports customer acquisition through dynamic conversion modeling. Chapter 2 turns to the back-end, analyzing how machine learning enhances customer retention by identifying and preventing churn. Chapter 4 extends the analysis to the layer of AI integration, investigating how organizations develop and govern the capabilities that enable such applications to function reliably over time. Rather than corresponding to discrete marketing stages, these chapters represent interdependent mechanisms within an evolving AI-enabled ecosystem—linking customer-facing applications with the organizational infrastructures that sustain them.

Analytical, Operational, and Organizational Constraints

The second contribution of this dissertation lies in identifying and examining the challenges that constrain AI integration in B2B settings. Across the three studies, these challenges emerge at multiple

levels—technical, analytical, managerial, and organizational—revealing the conditions under which AI's analytical potential encounters limits in real-world environments.

Within the machine learning and B2B analytics stream, Chapter 2 also advances understanding of the analytical and operational constraints that shape AI integration in business environments. Prior research has recognized the technical difficulty of incorporating heterogeneous and high-dimensional behavioral data into learning architectures, yet it offers limited guidance on how such data can be structured to mitigate sparsity, inconsistency, and preprocessing complexity (Coussement and Van den Poel, 2008). This study contributes a framework for structuring usage data along three analytical dimensions thereby transforming an underexplored and technically challenging data source into one that can be effectively operationalized for machine learning. In addition to addressing analytical tractability, this chapter introduces a complementary operational perspective by quantifying the carbon footprint and computational costs associated with model training and evaluation. This analysis exposes sustainability as a critical, yet often overlooked, constraint in large-scale AI experimentation. Collectively, these contributions reframe AI integration as a process that must balance analytical rigor with resource efficiency and environmental responsibility.

Chapter 3 also contributes to the AI and predictive-analytics methodology stream, particularly on the trade-off between model sophistication and managerial interpretability in B2B decision-making (Meire et al., 2017; González-Flores et al., 2025). Prior studies have emphasized the predictive benefits of complex models, yet their outputs are not consistently understood in organizational practice, especially when decisions depend on the trust and comprehension of non-technical users. This lack of interpretability constitutes a theoretical and operational gap, limiting AI integration despite technical performance gains. Chapter 3 addresses this gap by transforming model outputs into interpretable engagement trajectories that capture usage patterns across prospects, linking predictive outputs to recognizable prospect profiles. This contribution advances the literature by demonstrating a pathway to maintain accuracy while enhancing comprehensibility, analytical and managerial perspectives. In addition, the study provides managers with empirically derived usage profiles that facilitate evidence-based prioritization in targeting campaigns, enabling sales and marketing teams to operationalize AI-generated insights with greater confidence and accountability.

Chapter 4 contributes to the literature on AI challenges and governance in organizations, which examines the barriers that hinder scalable and reliable AI integration (Mikalef and Gupta, 2021; Weber et al., 2023; Schneider et al., 2023). Prior studies highlight issues such as limited data quality, methodological opacity, and insufficient cross-functional collaboration but tend to analyze these constraints in isolation, neglecting how they co-evolve and interact during the AI development process. This study addresses that gap by demonstrating that analytical, operational, and organizational constraints are interdependent and cumulative, emerging sequentially across the phases of data understanding and preparation, model definition, and tool integration. Using the concept of mirroring-hypothesis, the research shows that AI development reproduces existing organizational structures and communication patterns, making coordination breakdowns and governance misalignments central obstacles to integration. In doing so, the study reframes AI integration as a processual and systemic challenge rather than a set of discrete technical problems. For practitioners, these findings highlight that overcoming AI constraints requires continuous cross-functional coordination, adaptive governance mechanisms, and structural flexibility to manage the interdependencies between technical design and organizational practice.

Embedding and Sustaining AI in Organizational Practice

The third contribution centers on how AI systems are implemented and sustained in organizational practice. This contribution moves beyond model development or identification of barriers to examine

the mechanisms through which AI becomes operational, routinized, and continuously valuable within firms. Chapter 4 advances the organizational design and AI co-creation literature, which explores how technological and organizational systems evolve together during AI implementation (Bailey and Barley, 2020; Lebovitz et al., 2021; Van den Broek et al., 2021; Waardenburg and Huysman, 2022). Existing research has primarily treated mirroring between technology and organization as a static relationship (Colfer and Baldwin, 2016; Burton and Galvin, 2022), offering limited insight into how this relationship unfolds through time or with multiple stakeholders. Addressing this theoretical gap, the study provides a processual account of mirroring, showing how technological and organizational requirements coevolve across three phases. The findings reveal that the mirroring process is iterative with organizational assumptions and limited end-user participation often become inscribed into the final tool, reducing its usability and alignment with work practices. This insight reframes AI embedding as a co-creation process shaped by co-creation, learning, and adaptation. From a managerial perspective, sustaining AI in practice demands the cultivation of dynamic coordination capabilities, early and balanced stakeholder involvement, and continuous feedback loops that align the tool's functionality with evolving user needs and organizational objectives.

Chapter 3 further contributes to the AI implementation and data-driven marketing literature by demonstrating how predictive analytics can be operationalized to guide real-time decision-making and resource allocation in B2B contexts (Blattberg et al., 2008; Li et al., 2016). Whereas existing research on AI-enabled lead management has primarily emphasized predictive accuracy, this study advances understanding of how predictive outputs translate into economic impact through simulation-based decision analysis. By integrating weekly model forecasts with a profit simulation framework, the study quantifies the expected financial returns of alternative outreach strategies, revealing that model-guided interventions substantially increase conversion-related revenues while reducing inefficient prospect engagement. This approach positions AI implementation as a process of continuous value calibration, where prediction, action, and feedback are systematically linked. The study bridges the gap between algorithmic prediction and managerial control by embedding economic evaluation within the AI workflow. For practitioners, the findings highlight that realizing AI's financial potential requires the alignment of analytical processes with operational cadence, ensuring that model-guided actions are synchronized with sales planning, campaign timing, and resource constraints to maximize return on marketing investment.

References

Agarwal, R. and Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.

Bahoo, S., Cucculelli, M., Goga, X., and Mondolo, J. (2024). Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis.

Bailey, D. E. and Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, 30(2).

Baldwin, C. Y. (2023). Design rules: past and future. *Industrial and Corporate Change*, 32(1):11–27.

Baldwin, C. Y. and Clark, K. B. (2000). *Design Rules*, volume 1. The MIT Press.

Blattberg, R. C., Byung-Do, K., and Neslin, S. A. (2008). *Database Marketing*. Springer.

Burton, N. and Galvin, P. (2022). Modularity, value and exceptions to the mirroring hypothesis. *Journal of Business Research*, 151:635–650.

Campagnolo, D. and Camuffo, A. (2010). The concept of modularity in management studies: A literature review.

Chatterjee, S., Chaudhuri, R., and Vrontis, D. (2022). AI and digitalization in relationship management: Impact of adopting AI-embedded CRM system. *Journal of Business Research*, 150:437–450.

Chatterjee, S., Mikalef, P., Khorana, S., and Kizgin, H. (2024). Assessing the Implementation of AI Integrated CRM System for B2C Relationship Management: Integrating Contingency Theory and Dynamic Capability View Theory. *Information Systems Frontiers*, 26(3):967–985.

Chen, G., Huang, L., Xiao, S., Zhang, C., and Zhao, H. (2024). Attending to Customer Attention: A Novel Deep Learning Method for Leveraging Multimodal Online Reviews to Enhance Sales Prediction. *Information Systems Research*, 35(2):829–849.

Chen, K., Hu, Y. H., and Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and e-Business Management*, 13(3):475–494.

Colfer, L. J. and Baldwin, C. Y. (2016). The mirroring hypothesis: theory, evidence, and exceptions. *Industrial and Corporate Change*, 25(5):709–738.

Coussement, K. and Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1):313–327.

Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198. Association for Computing Machinery, Inc.

Dubey, R., Bryde, D. J., Blome, C., Roubaud, D., and Giannakis, M. (2021). Facilitating artificial intelligence powered supply chain analytics through alliance management during the pandemic crises in the B2B context. *Industrial Marketing Management*, 96:135–146.

Dwivedi, Y. K. and Wang, Y. (2022). Guest editorial: Artificial intelligence for B2B marketing: Challenges and opportunities.

Enholm, I. M., Papagiannidis, E., Mikalef, P., and Krogstie, J. (2022). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, 24:1709–1734.

Fildes, R., Ma, S., and Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4):1283–1318.

Gattermann-Itschert, T. and Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*, 107:134–147.

George, G., Haas, M. R., and Pentland, A. (2014). From the editors: Big data and management.

Gomez-Uribe, C. A. and Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19.

González-Flores, L., Rubiano-Moreno, J., and Sosa-Gómez, G. (2025). The relevance of lead prioritization: a B2B lead scoring model based on machine learning. *Frontiers in Artificial Intelligence*, 8.

Gordini, N. and Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62:100–107.

Grover, V., Chiang, R. H. L., Liang, T.-P., and Zhang, D. (2018). Creating Strategic Business Value from Big Data Analytics. *Journal of Management Information Systems*, 35(2):388–423.

Jahromi, A. T., Stakhovych, S., and Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7):1258–1268.

Janssens, B., Bogaert, M., Bagué, A., and Van den Poel, D. (2022). B2Boost: instance-dependent profit-driven modelling of B2B churn. *Annals of Operations Research*.

Järvinen, J. and Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management*, 54:164–175.

Jordan, M. I. and Mitchell, T. M. (2015). Data, privacy, and the greater good. *Science*, 349(6245):255–260.

Keegan, B. J., Dennehy, D., and Naudé, P. (2024). Implementing Artificial Intelligence in Traditional B2B Marketing Practices: An Activity Theory Perspective. *Information Systems Frontiers*, 26(3):1025–1039.

Kellogg, K. C., Valentine, M. A., and Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1):366–410.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6):84–90.

Lebovitz, S., Levina, N., and Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly: Management Information Systems*, 45(3):1501–1525.

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Ledro, C., Nosella, A., Vinelli, A., Dalla Pozza, I., and Souverain, T. (2025). Artificial intelligence in customer relationship management: A systematic framework for a successful integration. *Journal of Business Research*, 199.

Li, D., Chen, X., Becchi, M., and Zong, Z. (2016). Evaluating the energy efficiency of deep convolutional neural networks on CPUs and GPUs. In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 477–484. IEEE.

Li, H. (2022). Converting free users to paid subscribers in the SaaS context: The impact of marketing touchpoints, message content, and usage. *Production and Operations Management*, 31(5):2185–2203.

Lilien, G. L. (2016). The B2B Knowledge Gap. *International Journal of Research in Marketing*, 33(3):543–556.

Lim, W. M., Ahmed, P. K., and Ali, M. Y. (2019). Data and resource maximization in business-to-business marketing experiments: Methodological insights from data partitioning. *Industrial Marketing Management*, 76:136–143.

Linden, G., Smith, B., and York, J. (2017). Two Decades of Recommender Systems at Amazon. *IEEE Internet Computing*, 7(1):12–18.

Ma, T., Hu, Y., Lu, Y., and Bhattacharyya, S. (2024). Customer Engagement Prediction on Social Media: A Graph Neural Network Method. *Information Systems Research*.

Mariani, M. M., Machado, I., and Nambisan, S. (2023). Types of innovation and artificial intelligence: A systematic quantitative literature review and research agenda. *Journal of Business Research*, 155.

Mcafee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard business review*, 90:60–68.

Meire, M., Ballings, M., and Van den Poel, D. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. *Decision Support Systems*, 104:26–37.

Mikalef, P. and Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information and Management*, 58(3).

Mojir, N. and Anbil, S. (2025). The Value of Professional Ties in B2B Markets. *Marketing Science*, 44(5):1058–1081.

Moradi, M. and Dass, M. (2022). Applications of artificial intelligence in B2B marketing: Challenges and future directions. *Industrial Marketing Management*, 107:300–314.

Morgan, R. M. and Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, 58(3):20–38.

Nyadzayo, M. W., Casidy, R., and Thaichon, P. (2020). B2B purchase engagement: Examining the key drivers and outcomes in professional services. *Industrial Marketing Management*, 85:197–208.

Paschen, J., Kietzmann, J., and Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business and Industrial Marketing*, 34(7):1410–1419.

Puranam, P., Raveendran, M., and Knudsen, T. (2012). Organization design: The epistemic interdependence perspective.

Purmonen, A., Jaakkola, E., and Terho, H. (2023). B2B customer journeys: Conceptualization and an integrative framework. *Industrial Marketing Management*, 113:74–87.

Ravasi, D. and Stigliani, I. (2012). Product Design: A Review and Research Agenda for Management Studies. *International Journal of Management Reviews*, 14(4):464–488.

Schneider, J., Abraham, R., Meske, C., and Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses. *Information Systems Management*, 40(3):229–249.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., and Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. *Advances in neural information processing systems*, 28.

Shrestha, Y. R., Krishna, V., and von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, 123:588–603.

Sjödin, D., Parida, V., Palmié, M., and Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, 134:574–587.

Syam, N. and Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, 69:135–146.

Van den Broek, E., Sergeeva, A., and Huysman, M. (2021). When the machine meets the expert: An ethnography of developing ai for hiring. *MIS Quarterly: Management Information Systems*, 45(3):1557–1580.

Vlačić, B., Corbo, L., Costa e Silva, S., and Dabić, M. (2021). The evolving role of artificial intelligence in marketing: A review and research agenda.

Von Krogh, G. (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, 4(4):404–409.

Waardenburg, L. and Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, 32(4).

Wang, T., He, C., Jin, F., and Hu, Y. J. (2022). Evaluating the Effectiveness of Marketing Campaigns for Malls Using a Novel Interpretable Machine Learning Model. *Information Systems Research*, 33(2):659–677.

Weber, M., Engert, M., Schaffer, N., Weking, J., and Krcmar, H. (2023). Organizational Capabilities for AI Implementation—Coping with Inscrutability and Data Dependency in AI. *Information Systems Frontiers*, 25(4):1549–1569.

Wiersema, F. (2013). The B2B Agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, 42(4):470–488.

Yoganarasimhan, H., Barzegary, E., and Pani, A. (2023). Design and Evaluation of Optimal Free Trials. *Management Science*, 69(6):3220–3240.

Chapter 2

Incorporating Usage Data for B2B Churn Prediction Modeling

Chapter 2
Incorporating Usage Data for B2B Churn Prediction Modeling

Abstract

The potential for predicting churn in B2B contexts remains untapped despite the growing availability of digital customer traces, particularly usage data. Nevertheless, this source provides valuable insights into customer behavior and product interaction, serving as a powerful tool for understanding customer needs and improving retention efforts. This paper aims to explore the value of usage data for B2B customer churn prediction using a real-world dataset with 3,959 subscriptions from a European software provider. The contribution to literature is two-fold. First, we define a framework to structure usage data into cross-sectional features that could be included in predictive models. We study the impact of usage information on churn prediction performance along three primary components– timing, granularity, and expertise. The findings confirm the value of this data source and two of the three examined components, particularly in terms of AUC and TDL. Second, we provide insights into the carbon footprint of the entire modeling process required to integrate usage data into different machine-learning algorithms. Furthermore, to strengthen the robustness of our findings, we compare five common machine-learning algorithms in CCP. Consequently, this study uncovered unexplored aspects of usage data and practical implications for enhancing churn prediction in a B2B setting.

Keywords: B2B customer churn, Churn prediction, Usage data, Machine learning, Carbon footprint

2.1 Introduction

Business-to-business (B2B) firms establish unique relationships with their customers. These relationships impact the decision-making purchase process, which can be notably influenced by factors such as costs and budgetary considerations (Webster and Wind, 1972). Compared to Business-to-consumer (B2C), purchase decisions in B2B settings are more complex and often involve dynamic and heterogeneous constellations of stakeholders (Lievens and Blažević, 2021) and high-value transactions (Rauyruen and Miller, 2007; Jahromi et al., 2014; De Caigny et al., 2021). As a result, companies must establish long-term and reliable relationships with their customers and formulate effective and tailored retention strategies to encourage their future purchases and advocacy (Rauyruen and Miller, 2007). Therefore, firms invest substantially in customer retention efforts as a part of their customer relationship management strategies. The most effective customer retention management systems are proactive, allowing firms to take action before customers churn. Such systems require firms to identify clients at high risk of churn (Gattermann-Itschert and Thonemann, 2022) and take tailored retention actions based on this information. It is crucial to consider the distinctions of the B2B framework, which continue to be of utmost importance given the potential impact of retention initiatives on financial results (Janssens et al., 2022) due to the high value of industrial customers (Rauyruen and Miller, 2007; Jahromi et al., 2014; De Caigny et al., 2021). Hence, customer churn prediction (CCP) emerges as a powerful and efficient tool for implementing customer retention strategies. This presents an opportunity worth investigating in B2B contexts, considering the potential rewards for suppliers who effectively create and sustain loyal customer relationships (Rauyruen and Miller, 2007).

Data-driven B2B marketing has received significantly less research attention than its B2C counterpart for several reasons such as the complexity and heterogeneity of the problem domain, limited availability of data, and missing domain knowledge among researchers (Lilien, 2016). Consequently, research on CCP in B2B settings remains untapped despite the increasing popularity of CCP models in other contexts and the relevance of customer retention management for B2B companies. Previous studies developed CCP models in B2B contexts by testing the predictive performance of various machine-learning algorithms, including traditional approaches such as Logistic Regression (LR) or Decision Trees (DT) (Barfar et al., 2017; Chen et al., 2015); more complex algorithms such as Random Forest (RF), Support vector machine (SVM) (Gattermann-Itschert and Thonemann, 2022), or XGBoost (XGB); and some variations including cost or profit components (Janssens et al., 2022).

To improve the predictive performance of CCP models, previous research has acknowledged the importance of exploring innovative data sources and new types of features, including variables related to products and customer profiles (Jahromi et al., 2014; Gordini and Veglio, 2017). A relevant data source for assessing customer health is *usage data* (Hochstein et al., 2023). This data, encompassing digital user interactions with a product or service, offers valuable insights into usage patterns and facilitates the quantification of customer value realization. By generating metrics to assess the effectiveness of meeting customer needs, it provides a thorough understanding of customer engagement and satisfaction levels. By understanding changes in

usage, companies can identify at-risk customers who may need prompt intervention, highlighting the anticipated relevance of usage data in CCP models.

Cases such as John Deere's transition from conventional product offerings to subscription-based "as-a-service" models offering crop management services exemplify a shift in B2B product sales toward continuous digital solutions (Hochstein et al., 2023). The emergence of new market offerings, such as personalized services and digital experiences, has contributed to the increasing prevalence of data about how customers interact and use the offered products or services. B2B organizations have the potential to capture a wide range of digital usage data, such as usage of telecommunication services (Somosi et al., 2021) or clickstream data (Hochstein et al., 2023). Interestingly, despite evidence highlighting its value in different settings (Zhang et al., 2016; Nevskaya and Albuquerque, 2019; Zdravevski et al., 2020), research on the integration of usage data into CCP models in the B2B context is lacking.

Previous research in CCP for B2B contexts leveraged a variety of data sources, such as *firmographics* explored by Barfar et al. (2017), the value of *customer transactions* proven by Jahromi et al. (2014) and Gattermann-Itschert and Thonemann (2022), or *support* information for capturing undesired events (Janssens et al., 2022). Despite this, previous research has remained limited in exploring the potential of novel digital data sources, such as customer usage information, this gap in knowledge limits how firms can learn from usage experiences (Moradi and Dass, 2022). However, integrating this data source into predictive models presents challenges due to its high dimensionality, time variance, and sparcity. Our research contributes to the literature by introducing a novel framework that exploits the potential of usage data in CCP, particularly for the B2B setting. This novel framework incorporates three components: timing, granularity, and expertise, allowing for the integration of usage data within CCP models and leading to improved prediction accuracy. Additionally, this framework facilitates the inclusion of contextually nuanced and industry-specific factors, enhancing its adaptability to diverse business scenarios.

Previous studies have compared the value of CCP models in B2B contexts by calculating different sets of metrics. All the studies presented in Table 2.1 calculate the area under the curve ROC (AUC), except for Chen et al. (2015). Additional metrics related to predictive performance have been included, such as top decile lift (Gordini and Veglio (2017), Gattermann-Itschert and Thonemann (2022)), accuracy, or precision (Chen et al., 2015). Alternatively, Janssens et al. (2022) incorporated a different type of metric related to profit in retention campaigns(EMPB), particularly for B2B settings. While quantifying and understanding the predictive and profitable value of usage data in CCP models is relevant, it's important to consider the challenges associated with data collection and preprocessing, such as the carbon footprint. Therefore, our research contributes by incorporating an additional metric, the quantification of the carbon footprint. This comprehensive analysis goes beyond traditional metrics, allowing us to assess the model's predictive capability, expected profit in retention campaigns, and environmental footprint.

We validate our results using real-world data from a leading European software service provider. Our methodology for evaluating churn prediction algorithms, analyzing usage data, and assessing carbon footprint involves four main steps. First, we compare five common machine-learning algorithms and measure their predictive and profit performance using three CCP literature-based

metrics. Second, we quantify the significance of the usage variables in churn prediction. Third, we structure the usage data into cross-sectional features and explore their incorporation into churn prediction models. This approach considers timing, granularity, and expertise and provides insights into churn management strategies. Finally, to measure the trade-off between the value of the usage data and its carbon footprint, we assess the energy consumption of the entire modeling process. Our findings demonstrate that incorporating usage information improves customer behavior comprehension, improving churner identification and customized retention strategies. Additionally, our framework improves the performance of traditional CCP models that solely rely on *firmographic, transactional, or support* features.

The remainder of this paper is organized as follows. First, we review the literature on the prediction of B2B churn, including the data sources explored. Section 2.3 explains the proposed framework for structuring usage data into predictive variables for the CCP. Section 2.4 includes a detailed account of the case study and details of the predictive models and evaluation metrics. Section 2.5 presents the results of the study. Finally, Section 2.6 outlines the theoretical and managerial implications, limitations, and suggestions for further research.

2.2 Background

2.2.1 Customer retention in the B2B setting

The B2B context raises challenges regarding customer relationship management, given the importance of fostering strong and enduring relationships within B2B frameworks. This context demands that businesses prioritize strategies that promote customer retention and actively engage in relationship-building to ensure long-term profitability. Hence, previous research has mainly focused on identifying customers at risk of churning. These efforts have involved the inclusion of diverse data sources, modeling techniques, and the comparison of models using evaluation metrics based on predictive, domain-specific, and profit-related viewpoints.

					Data			
Year	**Paper**	**Algorithms**	**Metrics**	**Industry group**	**Firmographics**	**Transactional**	**Support**	**Usage**
2014	Jahromi et al. (2014)	Boosting, DT, Cost sens. DT LR	AUC, Cumulative Lift	Consumer Staples Distribution & Retail	✓			
2014	Chen et al. (2015)	NN (MLP), SVM, LR, C4.5 DT	Accuracy, Precision, Recall, F1	Transportation	✓	✓	✓	
2017	Gordini and Veglio (2017)	SVM(auc), LR, NN, SVM(acc)	AUC, TDL, Accuracy	Consumer Staples Distribution & Retail	✓	✓		
2017	Barfar et al. (2017)	LR, DT	AUC	Service provider Not specified)	✓		✓	
2022	Gattermann-Itschert and Thonemann (2022)	LR, SVM, RF	AUC, TDL, AP	Consumer Staples Distribution & Retail	✓	✓	✓	
2022	Janssens et al. (2022)	B2Boost (Incl. profit), VerbrakenBoost, XGBoost, RF Proflogit, Proftree, Lasso regression	AUC, EMPC, EMPB	Consumer Staples Distribution & Retail	✓	✓	✓	
	THIS STUDY	**LR, RF, XGBoost, DT**	**AUC, TDL, EMPB, Energy Consumption**	Software and services	✓	✓	✓	✓

Table 2.1: Previous churn prediction studies in B2B settings.
Note: *Industry classification based on Global Industry Classification Standard (GICS). Notation is included in Appendix A*

Table 2.1 summarizes a review of churn prediction studies in the B2B framework from

peer-reviewed ABS journals, ensuring the quality of the benchmark aligned with research from the field. It compares algorithm types, metrics, industries, and data sources included. We compare these attributes to identify the most accurate algorithms for churn prediction, the different churn drivers across industries, and the similarities or variations in the data sources used for churn prediction in the B2B framework. Gattermann-Itschert and Thonemann (2022) developed a churn prediction model for a wholesale setting, subsequently validated in a field study. Janssens et al. (2022) presented a novel maximum profit measure for B2B customer churn and integrated this metric into a gradient boosting classifier. Barfar et al. (2017) examined the link between service quality and churn, defining rationality and bounded rationality as two main assumptions influencing churn decisions. Gordini and Veglio (2017) tested the predictive capacity of Support Vector machines (SVM) by performing the hyperparameter setting based on the area under the curve ROC (AUC). The study demonstrates how the development of customer loyalty strategies outperforms commonly used management heuristics in the B2B e-commerce industry. Jahromi et al. (2014) compared different modeling techniques to predict customer attrition and developed a profit-maximizing campaign for customer retention. Finally, in the context of a logistics company Chen et al. (2015) studied the effect of length, recency, frequency, monetary value, and profit (LRFMP) on customer value to predict customer churn, expanding the original recency, frequency, and monetary value (RFM) approach with insights and managerial implications.

CCP models are classification algorithms that predict whether a customer will churn during a given observation period using a binary variable as a target (Coussement and De Bock, 2013). These models use input variables to estimate churn probability based on past information. Several algorithms have traditionally been implemented for CCP problems, as presented in Table 2.1. Based on a review of the literature on B2B CCP, the most popular algorithms are LR, DT, RF, XGB, and SVM. In addition, except for Gordini and Veglio (2017), all studies explored tree-based algorithms such as simple DT, RF, or variations in boosting. These algorithms exhibited the best performance along with LR and were implemented in the present study, as explained in Section 2.4.2.

To evaluate and compare the CCP models, validating their performance on unseen data is crucial to ensure a robust generalization of the models (Coussement and Van den Poel, 2008). It has been standard practice in previous studies to use multiple performance metrics when comparing churn prediction models to assess them from multiple perspectives. Except for Chen et al. (2015), all the studies presented in Table 2.1 include the AUC as a threshold-independent measure, together with accuracy, precision, lift, or profit-related measures. Section 2.4.3 explains the metrics used to evaluate the models examined in this study.

Finally, to address the real-world challenges faced by B2B marketers, it is essential to explore the research streams introduced by Mora Cortez and Johnston (2017). In particular, the industry context is relevant, given how dynamic the business environment can be and how the sector shapes elements influencing customer sales. In previous studies, the industries explored included wholesale, retail, e-commerce, and service providers (see Table 2.1). No prior research addressed CCP in a B2B context for business service providers. This industry's common subscription-based

revenue model requires careful customer retention efforts owing to small and stable cash flows, as maximizing profit requires avoiding churn. Previous research has highlighted how incorporating new data sources enhances the predictive performance of the CCP model (Janssens et al., 2022). Usage patterns significantly affect subscription renewals, emphasizing the importance of managing continuous usage intent for business service providers' success (Martins et al., 2019).

2.2.2 Data sources in CCP

A key angle to discuss is the integration of different data sources for CCP modeling. Firms generate and maintain large volumes of data in various formats from internal sources (Wang and Wang, 2020), such as customer transaction records, sales records of products, and customer feedback. Furthermore, the revolution in Big Data has led to the collection of abundant and diverse data on consumer behavior in real time, which, combined with information from various sources, generates novel insights that help marketers and researchers realize new gaps in the current understanding of customers (Erevelles et al., 2016). As mentioned by Shaw et al. (2001), the information and insights gained by applying decision support systems and data mining techniques may enhance existing knowledge within organizations.

For CCP, firms rely on easily accessible specific data sources. The first source is *firmographics*, which contains data related to firm characteristics, including company size, location, and industry. These data are often collected during the acquisition stage, making them an interesting source for churn prediction. In the B2B framework, this data source was included in most studies presented in Table 2.1. Barfar et al. (2017) expounded upon the significance of such variables as indicative measures of customer value, information that might be relevant for predicting customer churn.

Another common data source available to companies is *transactional data*. This information can be collected from multiple channels through which a company interacts with customers (Blattberg et al., 2008), and includes information about purchase history and expenditures. These data are commonly represented using RFM frameworks. In the B2B context, except for Barfar et al. (2017), all studies presented in Table 2.1 include this data source. Jahromi et al. (2014) validate the importance of frequency and recency, while Gattermann-Itschert and Thonemann (2022) highlights the critical role of monetary value variables for that particular case of study.

Finally, an additional data source commonly used in CCP studies in the B2B context is information on the *support* provided by the company regarding the quality of the product or service. Barfar et al. (2017) demonstrates the importance of monitoring every service failure to keep service pain low and minimize customer attrition for a service provider. Janssens et al. (2022) further elaborates on how interactions, such as calls or emails expressing dissatisfaction, may signal undesired events. Based on this rationale, this study considers various customer interactions as *support* data, aligning with the importance of customer service quality dynamics in CCP in the B2B context. The integration of diverse data sources, as detailed in Table 2.1, proves relevant, offering a comprehensive understanding of customer behavior and needs.

In addition, finding additional sources of information that can capture aspects related to customer needs and customer-owned touchpoints can augment CCP applications. For instance, Ascarza and Hardie (2013) illustrates the relevance of integrating various data types into CCP

models and highlights the crucial role of usage data in a contractual B2C setting, in this case, customers reduce their usage as they approach the renewal point. In this context, purchases represent usage, as a subscription allows customers to buy products or services similar to a "warehouse club." Furthermore, they elaborated on how customers assessed their foreseen usage levels for renewal decisions by comparing usage expectations with the associated costs. The preceding examples support the idea of exploring further applications of usage data and where they can be valuable, complementing traditionally explored data sources.

Additional cases in which usage data are employed in CCP models for the B2C context can be found in Zhang et al. (2016) and Nevskaya and Albuquerque (2019), which include usage data from the telecommunications and gymnasium subscription industries. A more recent study by Zdravevski et al. (2020) focused on a video streaming service and made a technical contribution to automatic feature extraction from various sources. This study also provided relevant business insights by revealing that user activity in the two weeks before the prediction day was the most informative for predicting customer churn. Similarly, the study showed that users who were still active for more than two months after subscribing were less likely to cancel their subscriptions.

Moreover, the inclusion of usage data in B2B frameworks has proven fundamental for detecting at-risk customers. Somosi et al. (2021) studied the effect of service elimination in the telecommunications industry on customer behavior by including usage intensity information related to calls and data. The results showed that usage intensity, among other variables, reduced the effect of price increases on customer churn. Other sources, such as Forrester Consulting (2014), evaluated how customer health scores were measured and monitored in the B2B SaaS software space, which indicated customer satisfaction and loyalty. This study highlights organizations with effective customer success management initiatives that employ various measures and incorporate product usage as the primary indicator to monitor customer health. This approach aligns with Wiersema (2013), who remarks on the importance of understanding customer-product interactions and how usage patterns evolve over time, providing valuable insights into relevant strategies to decrease churn rates.

Lilien (2016) highlights that while customer data collection in B2B companies is of utmost importance, it is nonetheless crucial for companies to use this information effectively. Table 2.1 shows that usage data remain unexplored in B2B churn prediction. To bridge this research gap, we recognize the complexity of organizational decisions, which often involve dynamic and heterogeneous groups of stakeholders (Lievens and Blažević, 2021). This study incorporates usage features into CCP models in a B2B case study by structuring usage data based on a novel framework. We evaluate the impact of this new data source on the performance of various machine-learning algorithms and the relevant business insights derived from it.

2.3 Representing usage data

Value quantification in B2B relationships is an unsolved problem exposed by Mora Cortez and Johnston (2017), who highlighted the relevance of understanding usage patterns to capture the real value provided to customers. With the growing digitalization of the B2B sector, there is

increasing interest in revealing additional dimensions of decision-making dynamics within the B2B context (Zhang and Chang, 2021). Usage data hold valuable information that can answer questions about when, what, or how customers interact with the focal product or service. This study establishes a comprehensive framework for integrating usage data into CCP models to further investigate these questions, which can lead to relevant managerial insights.

Despite the challenges of high dimensionality, time variance, and sparsity, usage data offer valuable insights into customer interactions. This study establishes a framework to address these challenges and provides relevant managerial insights. Issues of dimensionality and time variance arise from capturing extensive data per user, which adds complexity, particularly to a large customer base. Sparsity, which is characterized by sporadic usage periods, creates gaps in the data, limiting comprehensive information for specific periods or activities.

The attributes described above highlight the divergence of usage data from traditional data sources outlined in Section 2.2.2. For instance, usage information is commonly stored as data records in standardized formats. As highlighted by Zdravevski et al. (2020), data aggregation techniques can effectively reduce the number of rows in computer-generated logs while retaining valuable information. Therefore, this study integrates the three relevant components of usage. First, the *timing* component allows data consolidation from various usage periods at each decision point. A similar approach was applied by Ascarza and Hardie (2013), where the analysis per period allowed for the exploration of latent variables based on usage and a segmentation model over time. In terms of *granularity*, session analysis allows for the decomposition of usage behavior. Studies such as Schweidel and Moe (2016) and Rafieian and Yoganarasimhan (2022) explored patterns within individual sessions and across multiple sessions related to advertising while using a broadcasting platform and an app. Regarding the *expertise* dimension, studies such as Shi and Zhang (2014) and Lakshmanan and Krishnan (2011) discussed the relevant impact of the usage experience and learning process on product perceptions and customer behaviors over time. These components shape the framework for creating predictive usage variables, facilitating the extraction of user experience information. These features can be adapted for integration into diverse contexts with the available usage information.

2.3.1 Timing

Timing plays a significant role in usage data, comprising the time-related information revealed when customers actively use a focal product or service. Various data mining techniques can be used to extract essential features based on predefined periods by analyzing granular data and selecting important pieces of information (Zdravevski et al., 2020). Timing features can provide valuable information on customer involvement, preferences, usage patterns, and the evolution of customer relationships with products over time. To ensure relevant findings, the timing component must be defined according to the nature of the data and the context. This approach entails selecting an optimal frequency to capture prevailing trends and patterns, facilitating the aggregation of usage data, and enabling comparisons of metrics across various time intervals.

2.3.2 Granularity

The aggregation level is a crucial aspect to consider when defining cross-sectional features based on usage data. A higher granularity can lead to a more accurate understanding of usage patterns, although it may involve increased complexity. Usage data are commonly stored in terms of sessions. A session can be defined as a particular episode beginning with an interaction between the user and a product or service and includes all subsequent actions performed until a final point. The duration of a session can be arbitrary and can last for minutes, hours, or days (Zdravevski et al., 2020). Hence, to integrate this component into the feature engineering process, having information that enables the tracking of the session start and end times for each user is crucial.

2.3.3 Expertise

As Mora Cortez and Johnston (2017) note, customer education and its role in the consumption experience are key to improving profitability in the interaction process. When referring to usage, an essential concept to consider is users' proficiency, as it directly affects their requirements and interactions with products. Users perform different types of tasks based on their expertise levels. Consequently, companies may benefit from a deep understanding of their customers, including specific user profiles and valued activities, when interacting with the focal product or service. In this context, while expertise is inherently a continuous concept, it can be represented as a categorical attribute. Moreover, integrating business knowledge is crucial for classifying interactions as low- or high-expertise tasks.

2.4 Empirical analysis

This study aims to assess the added value of a novel data source, specifically, usage data in CCP models within the B2B setting. Thus, the empirical analysis covers the description of the dataset and feature creation process. Second, an overview of five machine-learning algorithms is evaluated by comparing the churn prediction models. Third, the evaluation metrics employed in the benchmarks are described. Fourth, hyperparameter tuning and cross-validation procedures are used to train the predictive models.

2.4.1 Data

This study is based on data provided by a European software service provider. The firm offers subscriptions to its automation software to a diverse set of industrial customers through various sales channels. In this case, we focus on online renewable subscriptions at annual intervals. Each subscription contains information on firm graphics, transactional history, support information, and usage data, as listed in Table 2.1. This is true for contractual churn. Churners are customers who do not renew their subscriptions after the annual renewal point, including the 15-day grace period.

Category	Subcategory	No. variables
Firmographics (F)	Region	9
	Currency	2
Transactional (T)	No. of renewals	1
	Length of relationship	1
	Maintenance contracts	1
Support (S)	No. of training courses taken per status	13
	No. of courses taken per level	4
	No. of courses taken per product	5
	No. of courses taken per type	5
	Unique courses	1
	connection to training platform	1
	Ever had support cases?	1
	Contact via campaigns	2
	Visits to website	1
Usage (U)	No. of actions	30
	Time of usage	30
	No. of files processed	20
	Size of files processed	20
	Changes in usage	4

Table 2.2: Summary independent variables per categories

The focal dataset comprises 3,959 subscriptions over three years (April 2019 to July 2022), with a churn rate of 11.9%. Based on information provided by the company, the final dataset consists of 151 independent variables for each annual subscription. A one-year period is consistently defined to monitor information during the subscription interval.

As presented in Section 2.2.2, churn prediction models include different types of data sources, and the predictive variables are grouped into four categories: *firmographics, transactional, support, and usage*. Table 2.2 presents an overview of these categories and variables created after the feature engineering process.

The company has multiple data sources. *Firmographic* variables consider customer location and subscription payment currency. *Transactional* variables involve past renewals, first subscriptions, and additional product purchases. *Support* related data is drawn from a customer training platform offering e-learning courses, and marketing campaigns and support cases contribute to variables like campaign contacts and cases submitted during the subscription period. One of this study's main contributions is structuring the usage data into cross-sectional features that can be included in predictive models. In this context, usage data refer to tracking user interactions with software. This approach includes recording the date, time, actions performed, session identifiers, and additional details of the processed documents.

Figure 2.1 provides a graphical depiction of the experimental choices undertaken, aligned with the framework outlined in Section 2.3. Regarding the *timing* dimension, subscription year usage information was grouped into quarters (Q1, Q2, Q3, and Q4) to track usage patterns. This approach was implemented to aggregate information while monitoring the evolution of usage patterns throughout the subscription year. Regarding *granularity*, the data were initially consolidated based on usage sessions, utilizing a session identifier. Subsequently, the information was aggregated by calculating average usage values per session. This method provided a more

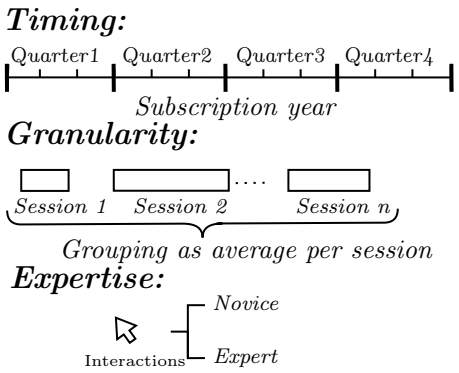


Figure 2.1: Experimental choices to create usage features

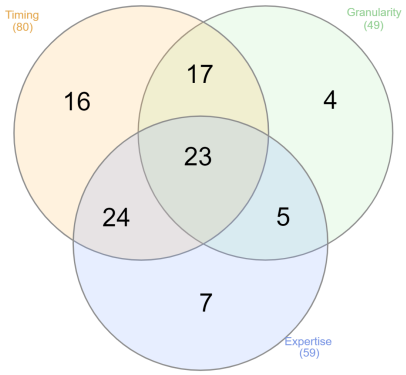


Figure 2.2: Usage variables per component (Heberle et al., 2015).
Note: *The figure does not include the eight general variables that exclude variations in the usage components (two for each usage value)*

nuanced understanding of user engagement within individual sessions. Concerning *expertise*, all available actions in the software were classified into two proficiency levels: novice and expert. These classifications were collaboratively established with company members with deeper business knowledge. The primary objective was to discern which actions required a higher level of proficiency. For instance, actions such as complex document manipulation or advanced feature utilization may be categorized as requiring expert-level proficiency.

Furthermore, this framework allows for the integration of contextually nuanced and industry-specific factors that contribute to the value derived from usage within a specific case study. For this purpose, it is crucial to understand the original input data and context of the products or services provided. In this framework, the service provided is an automation software used to process files, which allows users to perform different tasks according to their needs. The usage value provided by the software consists of three primary elements: the number of actions performed, duration of usage, and files processed in terms of size and volume. Figure 2.2 shows the number of usage features created by integrating the usage values and framework presented in Section 2.3 with the three usage components. Appendix B provides two examples to enhance the clarity of the created features, and Appendix C presents a summary of all the usage features created to clarify the different combinations of usage dimensions.

An essential step in the modeling process is data preparation. Regarding data integrity, the task of revising missing values was considered the nature of the variables. First, instances where renewals did not occur were substituted with a zero value, indicating that customers had not renewed their subscriptions. Furthermore, given the diversity of scales exhibited by different variables, a standardization procedure was applied to ensure that all scaled features resided in a similar numerical range (Crone et al., 2006). Finally, categorical features were transformed into dummy variables to make all the variables compatible with the different algorithms.

2.4.2 Classification algorithms

As presented in Table 2.1, *LR* is a prevalent benchmark frequently applied in churn prediction studies. This model assumes a linear relationship between the predictors and target, which makes it conceptually simple (Bucklin and Gupta, 1992) and easy to interpret. It also provides robust results compared with other classification techniques.

DT is a technique frequently used in churn prediction tasks and is characterized by a tree-shaped structure representing a set of decisions that establish rules for classifying a given dataset (Lee and Siau, 2001). The popularity of DTs as classifiers is largely owing to their transparency, interpretability, and ease of implementation (Olafsson et al., 2008). In this study, we aim to explore DTs because this algorithm has been widely used for churn prediction in B2B settings (Barfar et al., 2017; Gordini and Veglio, 2017; Jahromi et al., 2014; Chen et al., 2015). In addition, our research aims to examine the performance of both single and ensemble learners, similar to the approach of Jahromi et al. (2014). This comparative analysis provides insights into the advantages and limitations of each method, aiding in selecting the most appropriate approach for solving churn prediction problems.

The *RF* algorithm is an ensemble-based DTs. It creates multiple trees on the bootstrapped training samples in parallel. However, unlike in the standard DT algorithm, when a split is considered in a tree, only a subset of the total number of independent variables is used as a candidate predictor. Furthermore, in each split, a different set of predictors is used, increasing the variability of the trees and improving the reliability of the final predictions (Breiman, 2001). RFs were selected for implementation in this study because of their successful application in multiple CCP cases and predictive performance. Furthermore, RFs are known for their robustness to outliers and noise and ability to identify feature importance in the final model (Coussement and Van den Poel, 2008). However, certain hyperparameters must be set to implement this algorithm, such as the maximum depth of the trees, the number of features to consider when looking for the best split, and the number of trees to grow in the forest. Section 2.4.4 explains the hyperparameter tuning process in detail.

XGB is a boosting method that sequentially builds weak learners by creating DTs and calculating predictions using ensemble learning (Chen and Guestrin, 2016). The prediction error was calculated at each iteration, and the DTs were recreated to improve prediction accuracy. Boosting-based algorithms have demonstrated exceptional performance in diverse classification problems. However, the potential of CCP has not been extensively explored. In a recent study, Habel et al. (2023) developed a tool that estimates churn probabilities using XGBoost, demonstrating good performance for a B2B wholesaler.

SVM is an algorithm that creates a classification boundary and aims to maximize the distance between the classified instances to generalize new samples. In CCP problems SVM has been extensively used in B2C (Coussement and Van den Poel, 2008; Chen et al., 2012) and B2B settings (Gordini and Veglio, 2017; Gattermann-Itschert and Thonemann, 2022) demonstrating a strong performance in some cases.

2.4.3 Model evaluation

In this study, we used Area Under the ROC Curve (AUC) as a threshold-independent measure, TDL as a domain-specific measure (Gattermann-Itschert and Thonemann, 2022), and the Expected maximum profit measure ($EMPB$) as a profit-based measure (Janssens et al., 2022) to evaluate all modeling techniques based on the approaches observed in previous studies (see Table 2.1).

A common metric used in the field is AUC (Huang and Ling, 2005), which measures the area under the receiver operating characteristic (ROC) curve. This metric compares the rates of false and true positives, and the higher the AUC, the better the predictive power. The ROC curve is a graphical representation of the trade-off between a model's sensitivity (true positive rate) and specificity (false positive rate) and the precision-recall curve.

Another frequently adopted metric is the TDL, which measures the extent to which a model effectively identifies customers who are most likely to churn. This metric compares the predicted probability of churn for the top 10% of customers who churn in a group, and the resulting value is divided by the baseline churn rate (Xie et al., 2009). This metric helps companies prioritize their efforts and allocate resources more efficiently in retention strategies by targeting customers most likely to churn.

Considering some limitations of the previous metrics is vital because they assume that classification errors carry the same costs (Hand, 2009). In churn prediction, incorrectly classifying a churner as a non-churner results in a different cost than the opposite case; thus, it is crucial to consider this factor. Therefore, this study incorporated the EMPB metric, identifying a retention campaign's most profitable churn model. The EMPB metric requires the adjustment of multiple parameters as performance is measured based on total profitability, specifying the costs and benefits associated with a retention campaign. In terms of contacting cost, the value was set to 15 EUR, in line with Janssens et al. (2022), by considering the average cost of time of an account manager, who is more likely to execute a retention campaign in a B2B setting. Regarding the acceptance rate of the campaign, as this is an uncertain parameter that is difficult to determine, the definition proposed by Janssens et al. (2022) was followed using a β distribution with $\alpha = 6$ and $\beta = 14$, resulting in a final rate between 0.1 and 0.5. The incentive rate and Customes lifetime value (CLV) were defined based on Jahromi et al. (2014) and Janssens et al. (2022), the first with a value of 0.05 and the second based on purchases made in the last year of the independent period to capture differences in CLV due to the size of industrial customers.

2.4.4 Churn prediction models

To incorporate the elements presented in Sections 2.4.1, 2.4.2, and 2.4.3 and address the relative research gaps, we compared five different machine-learning algorithms, four sets of variables, and three types of evaluation metrics. This setup allowed us to compare the models' performances and quantify the impact of usage variables on churn prediction.

First, a predefined set of hyperparameter candidates was established to fine-tune diverse machine-learning algorithms and determine their optimal configurations. As shown in Table 2.3, the selection of different candidate settings was based on the B2B churn literature. To

Algorithm	Parameter	Specifications	Reference
Logistic regression	Penatly	L1, L2	Gattermann-Itschert and Thonemann (2022)
	Regularization	$[10^{-5}, 10^{-4}, ..., 10^{2}]$	
Decision Tree	Criterion	Gini, entropy	Chen et al. (2015)
	Min samples per leaf	15, 20, 25	
	Max features	$\sqrt{F}$, log2(F), all	
	Class weight	Balanced, None	
Random Forest	Max no. features	$\sqrt{F}, min(F, 2\sqrt{F}), min(F, 3\sqrt{F}), min(F, 4\sqrt{F})$	Gattermann-Itschert and Thonemann (2022)
	Max tree depth	[3, . . . ,. Min(F,15)]	
	Min samples per leaf	2	
XGBoosting	Boosting rounds	[2, 5, 10, 20, 50, 100, 200, 500]	Janssens et al. (2022)
	Learning rate	[0.001, 0.01, 0.1, 0.2, 0.5]	
	Gamma	[0.5, 1, 1.5, 2]	
SVM	Kernel	linear	Gattermann-Itschert and Thonemann (2022)
	Regularization C	[0.2,0.3,. . . ,1.2]	
	Class weight	balanced	

Table 2.3: Parameter setting

avoid overfitting, regularization, and penalties were explored by penalizing the model coefficients (Gattermann-Itschert and Thonemann, 2022). In terms of the DTs, the modified hyperparameters were the criterion, minimum number of samples per leaf, maximum number of features, and inclusion of weights for observations to deal with class imbalance (Chen et al., 2015). With respect to RF, the number of predictors and the maximum tree depth were varied for each set of features (Gattermann-Itschert and Thonemann, 2022). The XGB models also vary according to different boosting rounds, learning rates, and gamma parameters (Janssens et al., 2022). Table 2.3 summarizes the candidate settings selected for each algorithm.

The experimental design chosen to optimize the different hyperparameter settings for the algorithms consisted of a 5 × 3 cross-validation, where the dataset was split into three subsets: training, validation, and test. Each set contained one-third of the complete dataset (Burez and Van den Poel, 2009; De Caigny et al., 2020). The model was trained using the training dataset, and the resulting model was evaluated in the validation set to select the optimal combination of hyperparameters. The final model was retrained using the validation and training sets, and the final performance was assessed on a test set that contained unseen data. To ensure a proper training process, a pipeline consisting of three steps was implemented: first, the data were split into three subsets; second, preprocessing steps were performed (Coussement et al., 2017), and the training process was carried out.

To compare the results of the different algorithms and combinations of predictors to address the research gaps, an F-statistic was computed to evaluate the 5 × 3 cross-validation outcomes, as proposed by Alpaydin (1998) using formula 2.1. In this case, $p_i^{(j)}$ represents the difference between the performances of the two algorithms or two sets of features, according to the case of folds j = 1, 2, 3 of replication i = 1,..., 5.

$$f = \frac{\sum_{i=1}^{5} \sum_{j=1}^{3} (p_i^{(j)})^2}{3 \sum_{i=1}^{5} s_i^2} \tag{2.1}$$

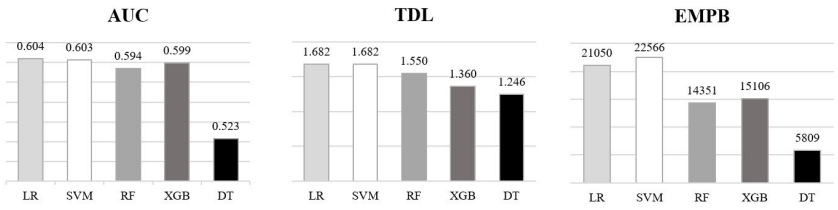


Figure 2.3: Evaluation metrics per algorithm

2.5 Results

This section presents the study's results and main findings. First, we present the performances of the various predictive models in terms of each evaluation metric (Section 2.4.3). Second, we analyze the global significance of usage features and assess the impact of different usage components. Third, relevant managerial implications are derived. Finally, a detailed analysis is presented to quantify the implications of incorporating cross-sectional usage features in churn prediction models, specifically in terms of carbon footprint.

2.5.1 Best algorithm

An initial model was constructed using the full set of predictors *(firmographics, transactional, support, and usage* (F_T_S_U)) to compare the five top-performing machine-learning algorithms in the literature for CCP models within the B2B domain. Each algorithm was fitted, as outlined in Section 2.4.4, to identify the optimal combination of hyperparameters for each evaluation metric. Figure 2.3 illustrates these results. LR outperforms the other algorithms in terms of TDL and EMPB. Regarding TDL, LR outperforms the other models by identifying 1.68 times more churners than random selection when contacting 10% of the customers with the highest probability of churn. Regarding the EMPB, the most profitable churn models in a retention campaign are LR and SVM, with an expected maximum profit above 21,000 EUR. In terms of AUC, LR, SVM, and XGB exhibit values close to 60%. Furthermore, the findings indicate that DT performs poorly on all three metrics. By contrast, RF and XGB exhibit comparable performances in accurately predicting churners and optimizing profits in a retention campaign. Table 2.4 presents the results of the pairwise comparisons. The analysis reveals significant differences in the performances of LR and DT for all metrics considered. Moreover, LR performs significantly better than XGB in terms of the TDL.

2.5.2 Added value and insights of usage data

Following a comparison of the results obtained from the different algorithms, the next task involved examining the added value of the usage data in the top-performing model. This study addresses the gap in the existing literature related to the inclusion of this valuable data source in B2B CCP models. As shown in Figure 2.4, this objective was accomplished through a three-stage process. First, we evaluated the performance of the models by incorporating different sets of available features to assess the incremental value derived from the inclusion of usage variables.

Algorithm 1	Algorithm 2	AUC		TDL		EMPB	
		F Test	Adj p-value	F Test	Adj p-value	F Test	Adj p-value
Logit	**Random Forest**	1.399	1.000	2.943	0.389	1.794	1.000
Logit	**XGBoost**	0.343	1.000	12.282	**0.012****	1.155	1.000
Logit	**DecisionTree**	105.762	**0.000*****	26.457	**0.000*****	11.324	**0.027****
Logit	**SVM**	0.033	1.000	0.000	1.000	0.061	1.000
Random Forest	**XGBoost**	0.522	1.000	5.145	0.187	0.026	1.000
Random Forest	**DecisionTree**	78.373	**0.000*****	15.978	**0.004*****	6.369	0.176
Random Forest	**SVM**	1.360	1.000	3.890	0.468	2.340	1.000
XGBoost	**DecisionTree**	117.260	**0.000*****	1.545	0.448	5.077	0.258
XGBoost	**SVM**	0.235	1.000	14.787	**0.008*****	1.617	1.000
DecisionTree	**SVM**	138.261	**0.000*****	33.053	**0.000*****	11.536	**0.041***

Notes: *The resulting p-values of the F-test are used to compare the performance of all algorithms in pairwise comparisons* Alpaydin (1998). *The adjusted p-values are calculated using* Holm (1979) *posthoc procedure. Significance levels of 90%, 95%, and 99% are represented by *, **, and ***, respectively, indicating significant differences.*

Table 2.4: Pairwise algorithm comparison

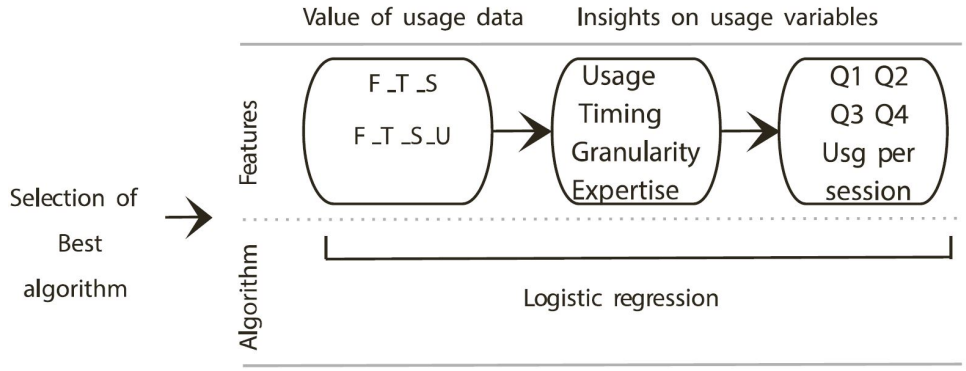


Figure 2.4: Added value of usage features

Subsequently, each usage component was analyzed by constructing separate churn prediction models incorporating the specific variables for each element. In the final phase, our primary focus was identifying the paramount usage components and their potential implications for the CCP.

Different models were trained and compared to appraise the value of the usage data, as shown in Table 2.5, which illustrates the model's performance when incorporating and excluding usage features. By predicting churn in both cases, the results demonstrate the additional value of incorporating usage variables, as evidenced by the superior performance of the three evaluation metrics compared to the other models. However, based on the F-test, the observed differences are statistically significant only for the AUC and TDL.

Performance				*Pairwise F - Test*					
Features	AUC	TDL	EMPB	AUC		TDL		EMPB	
				F Test	p-value	F Test	p-value	F Test	p-value
F_T_S	0.592	1.419	17,704.96	3.283	**0.097***	3.417	**0.090***	1.451	0.361
F_T_S_U	0.604	1.682	21,050.02						

Notes: *The resulting p-values of the F-test are used to compare the performance of both combinations of features* Alpaydin (1998). *Significance levels of 90%, 95%, and 99% are represented by *, **, and ***, respectively, indicating significant differences.*

Table 2.5: Evaluation metrics per blocks of features

Furthermore, we examined in detail the usage components presented in Section 2.3 to

understand the impact of each component on churn prediction and acquire a deeper comprehension of the business. To this end, usage variables were grouped into four categories. First, we considered variables that exclusively captured the temporal aspect, differentiating usage on a quarterly basis throughout the subscription year. Second, we included variables that provided a granular view of usage, specifically by calculating the average usage per session. Third, we incorporated features to assess the expertise level of usage based on the types of tasks performed. Finally, we introduced global features that aggregated the overall usage across the three defined components (i.e., total usage time during the subscription, total number of actions during the subscription, and total number of documents processed during the subscription). These features were included to assess the added value of each usage component.

Usage component	AUC	TDL	EMPB
Timing	**0.574 (0.016)**	**1.381 (0.279)**	14,204.42 (11,797,17)
Granularity	0.545 (0.023)	1.219 (0.165)	**15,902.88 (12,254.09)**
Expertise	0.524 (0.016)	1.046 (0.190)	10,340.94 (11,147.26)
Global features	0.527 (0.021)	1.122 (0.220)	10,463.72 (11,527.01)

Table 2.6: Evaluation metrics per usage component

Table 2.6 presents the separate evaluation of each usage component, excluding other types of variables to offer a clear comparison. According to the results, the timing features outperform the other usage components in terms of AUC and TDL, while the best results for EMPB are obtained via the granularity features. Table 2.7 shows the outcomes of pairwise comparison between the global usage features and each component. The study findings demonstrate that incorporating the timing component in the model can enhance its predictive performance and profitability, while the granularity features have a positive impact, particularly on profit. Due to the nature of the different performance metrics, we delved into a more detailed analysis of insights on the timing component in the subsequent results. This was done to gain a better understanding of patterns and potential churn risks specifically associated with AUC and TDL.

	AUC		TDL		EMPB	
	F Test	**Adj p-value**	**F Test**	**Adj p-value**	**F Test**	**Adj p-value**
Global features vs. Timing	45.207	**0.000*** **	7.910	**0.027** **	0.772	0.774
Global features vs. Granularity	5.067	**0.065***	1.871	0.365	1.568	0.663
Global features vs. Expertise	0.206	0.653	1.028	0.365	0.001	0.977

Notes: *The resulting p-values of the F-test are used to compare the performance all usage components in pairwise comparisons* Alpaydin (1998). *Significance levels of 90%, 95%, and 99% are represented by *, **, and ***, respectively, indicating significant differences.*

Table 2.7: Pairwise comparison usage blocks

To better understand the impact of usage on customer churn, we conducted an experimental training predictive model using two distinct feature sets: timing and granularity. Quarterly features, such as the number of actions, usage duration, and files processed, were compared, along with the average values per session. Table 2.8 summarizes the results for the average value per metric and the corresponding standard deviation. By examining the three metrics, the quarter-preceding renewal (Q4) usage information offers enhanced predictive and profitable power in assessing customer churn compared with the other timing blocks. Moreover, when

integrating the usage component of granularity, combining usage information as the average per session yields a higher impact on performance and profits than incorporating the total usage, as shown in Table 2.8.

Furthermore, Figure 2.5 presents the β-coefficients of four different models (one per quarter)and shows how the predicted probability is affected by the timing component, specifically focusing on its influence while excluding variations in expertise or granularity. It provides a clear depiction of this relationship across the four quarters and four usage features. It is interesting to observe how the positive and negative effects on churn vary over time for the four variables; notably, analyzing usage patterns throughout the subscription period is crucial as interactions with the software evolve. For instance, a higher usage duration increases the churn probability in the first two quarters but decreases it in the last two quarters. Similar effects are observed for processed documents, where a negative impact is noted throughout the year except in the third quarter. Consistent with the results in Table 2.8, it is crucial to understand the impact of usage behavior during Q4 on churn predictions. In this case, after some variations in the effects over the year, all features in Q4 have a negative impact on churn probability; however, the number of documents processed and the number of actions have the highest effect (coefficients of -0.19 and -0.18), followed by document size (-0.05), and usage time with the smallest impact (-0.02). Notably, heavy usage before renewal significantly decreases the churn probability, underscoring the importance of analyzing usage dynamics over time.

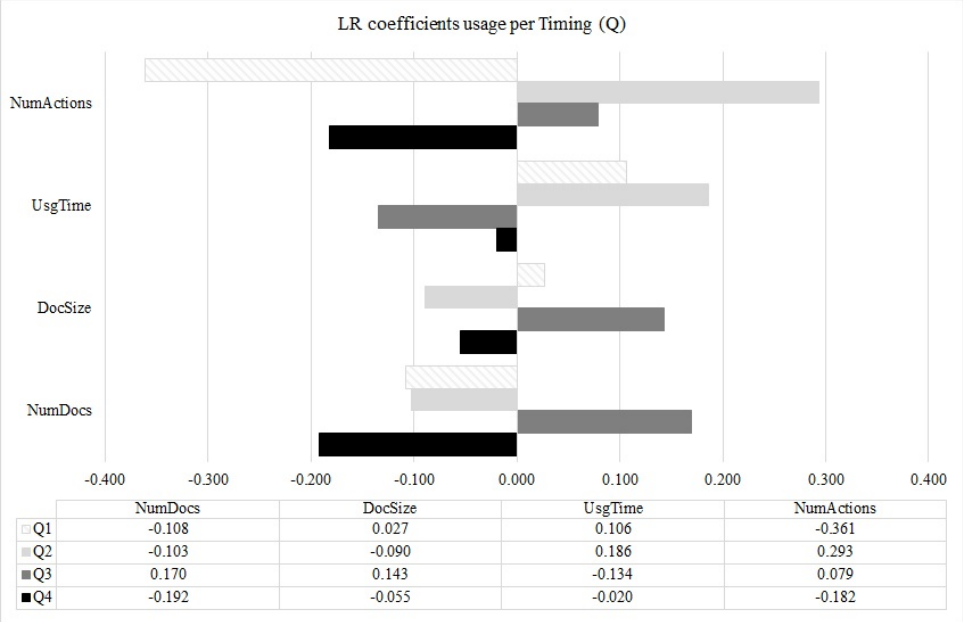
	Variables	AUC		TDL		EMPB	
Timing	**Quarter1**	**0.520**	**(0.030)**	**1.127**	**(0.306)**	**7,517.700**	**(7,838.662)**
	Quarter2	0.507	(0.014)	1.004	(0.210)	8,613.220	(13,145.233)
	Quarter3	0.506	(0.020)	0.856	(0.247)	6,882.460	(9,000.212)
	Quarter4	**0.573**	**(0.017)**	**1.441**	**(0.260)**	**14,002.860**	**(11,142.747)**
Granularity	**Session**	**0.535**	**(0.022)**	**1.266**	**(0.202)**	**12,152.380**	**(9,897.500)**
	Total	0.525	(0.022)	1.093	(0.240)	10,432.100	(10,842.267)

Table 2.8: Comparison timing and granularity components

2.5.3 Carbon footprint

We observe a trade-off between the value of information and its acquisition costs (Blattberg et al., 2008). Recent studies have explored the environmental costs in terms of the carbon footprint of training, developing, and implementing different types of predictive models. According to Li et al. (2016), as the size of a dataset increases exponentially, the energy demand for training such datasets increases rapidly. The assessment of the trade-off between the value and use of usage data in a predictive model is particularly pertinent owing to its high dimensionality. To further understand this trade-off, we approximated the carbon footprint associated with usage data during the feature creation and algorithm training processes by conducting a comparative analysis with the scenario excluding the usage data.

Previous studies have explored different methodologies for quantifying carbon footprint. First, García-Martín et al. (2019) presented an overview of the different approaches to estimating energy consumption in general and machine-learning applications in particular, with a detailed taxonomy,



	NumDocs	DocSize	UsgTime	NumActions
Q1	-0.108	0.027	0.106	-0.361
Q2	-0.103	-0.090	0.186	0.293
Q3	0.170	0.143	-0.134	0.079
Q4	-0.192	-0.055	-0.020	-0.182

Figure 2.5: Coefficients timing features

advantages, disadvantages, and possible machine-learning applications of the techniques. Later, Lacoste et al. (2019) presented a machine-learning emissions calculator, together with concepts and explanations of the factors affecting emissions, such as carbon intensity, computing infrastructure, and training time. Lottick et al. (2019) emphasized the need for computer scientists to actively engage in environmental sustainability by taking an honest and proactive role. The authors provided a Python package detailing the concepts, methodology, and data.

Henderson et al. (2020) and Anthony et al. (2020) provided tools for tracking real-time energy consumption and carbon emissions. The former also presented a leaderboard for energy-efficient reinforcement learning algorithms, and both studies proposed strategies for mitigating carbon emissions and reducing energy consumption. A more recent study by Meulemeester and Martens (2023) explored carbon footprint in terms of "common" data science tasks, in contrast to previous studies predominantly focused on large-scale advanced tasks. All the studies were intended to encourage researchers to incorporate power usage and estimate carbon emissions as additional metrics, encouraging a more comprehensive evaluation of the ecological impact associated with their research efforts.

Regarding the methodology employed, previous studies commonly encompassed a general definition for quantifying carbon footprint. This definition typically includes factors such as power consumption, which is directly linked to the specifications of the machine used and the processing time and conversion of average emissions per unit of energy related to the type of electricity generation. This approach was also used in the present study. We calculated the carbon footprint as follows: $PwrCons * Time * CO2eq/energy$. General assumptions were applied to derive the overarching power consumption and emission conversion results. First, considering that the models and experiments did not require special high-performance hardware, we used the power consumption reference of a server processor commonly used in data centers and enterprise applications, specifically AMD EPYC 7763. According to Bloomberg (2023),

		Excluding usg. Features		Including usg. Features		
		Carbon footprint (grams CO2 eq.)	Equivalence Km Driven by an average ICE car	Carbon footprint (grams CO2 eq.)	Equivalence Km Driven by an average ICE car	Percentage Increase
Feature creation		86.379	0.357	86.379	0.357	NA
Algorithm training	**LR**	1.270	0.005	2.349	0.010	84.96%
	DT	1.350	0.006	4.044	0.017	199.56%
	RF	7.119	0.029	21.290	0.088	199.06%
	XGB	18.330	0.076	79.877	0.330	335.77%
	SVM	27.340	0.113	40.990	0.169	49.93%

Table 2.9: Results Carbon footprint per stage

this processor typically requires 280 Watts under typical operating conditions. For emissions conversion, we employed the latest available reference values for the emissions intensity in Europe in 2021, provided by EMBER Climate (2023). These values were used to estimate the carbon dioxide emissions associated with energy consumption. The emission intensity value used is 296.96 grams of CO2 per kilowatt-hour (gCO2/kWh) of electricity consumed. Finally, to assess processing time, the corresponding experiments included code snippets to quantify the duration of each process. This approach allowed us to accurately measure and analyze the time spent during execution.

The results are presented in Table 2.9 in terms of grams of "CO2 equivalence". Nevertheless, to increase the interpretability of the results, they are converted to the quantity "km driven by an average ICE (internal combustion engine) car," in line with Anthony et al. (2020) and Meulemeester and Martens (2023). We used a reference value of 0.242 kgCO2eq. /km (EPA, 2023) to convert the grams of CO2eq. kilometers driven by a car. By dividing each result by this reference value, we obtained an approximate estimation of the distance in kilometers. As shown in Table 2.9, the feature creation process has the most significant impact on the carbon footprint. This task generates an equivalent amount of CO2 emissions, driving for approximately 0.357 km. This correlation may be attributed to the high dimensionality of the usage data and how the feature engineering process captures all valuable information within the defined features. Regarding the process of training different algorithms to determine the best-performing model, we observed how each algorithm dealt with an increase in the number of features. As presented in Table 2.9, for LR, the percentage increase in carbon footprint emissions is approximately 85%; for DT and RF, the increase corresponds to approximately 200%, for SVM the increase is close to 50%; and for XGB, the highest increase is 335%. It is important to note how the algorithm's complexity and the inclusion of new features play a vital role in processing time and, therefore, the carbon footprint.

By recognizing the interaction between algorithm complexity and data sources, researchers and practitioners can make informed decisions to mitigate the environmental impacts of data science tasks. This understanding encourages the exploration of algorithmic optimization and the conscientious selection of data sources to achieve more environmentally sustainable practices in the field of data science.

2.6 Discussion

2.6.1 Theoretical implications

Our research offers valuable contributions to the B2B CCP literature. First, we expand on the data sources studied in B2B churn prediction literature by focussing on usage data. We provide a transferable framework to structure usage data and evaluate the importance of each usage dimension. As such we extend earlier studies of CCP in the B2B domain, that focused on including firmographic, transactional, and support variables (Gattermann-Itschert and Thonemann (2022), Janssens et al. (2022), Barfar et al. (2017), Gordini and Veglio (2017), Jahromi et al. (2014), Chen et al. (2015)). Alternatively, in this research, we leverage frequently used variables and usage data to create a holistic view of B2B customers improving the performance of CCP models.

The proposed framework consists of three domain-independent components *(Timing, granularity, and expertise)*, facilitating the structuring and integration of usage data irrespective of the domain expertise, and promoting the accessibility of an underutilized data source. Besides, we empirically validated the proposed framework's effectiveness within a real-world B2B setting for a software and services company. This application demonstrates how firms can leverage AI to learn from user experience (Moradi and Dass, 2022). We achieve this by employing multiple predictive models and providing insights based on the results.

Second, in a B2B setting, we contribute to the literature on CCP models by developing a comprehensive understanding including the quantification of their performance in terms of carbon footprint together with profit and predictive power. While most previous B2B studies consider only predictive performance metrics, Janssens et al. (2022) included specific profit metrics for a retention campaign. Adding to this, in this study we include a new performance measure, the quantification of the carbon footprint. This comprehensive analysis allowed us to assess not only the model's ability to identify the risk of churn and target high-value customers but also its impact on the company's environmental footprint stemming from processing and integrating usage data into different machine-learning algorithms.

2.6.2 Managerial implications

From a managerial perspective, our research yields three valuable contributions. First, by understanding the added value of usage data (Moradi and Dass, 2022), businesses can better anticipate customer behaviors, tailor their marketing strategies, and enhance customer satisfaction. The findings provide an actionable outline for B2B companies seeking to improve their customer retention strategies, by leveraging available data sources to gain insights into customer needs and preferences. This case study reveals the company's relevant variables to predict customer attrition, such as usage when leading up to renewal and session aggregation.

Particularly, this study proposes to analyze the data per quarters, yet the definition of timing can be tailored for specific applications. However, as the findings demonstrate, evaluating usage patterns over various periods can notably improve the precision of churn prediction models. In addition, aggregating the usage information into sessions enables a more in-depth analysis. This

approach allows the identification of specific features frequently used across multiple sessions, facilitating the recognition of recurring usage patterns.

Second, by understanding the foreseen value of usage data in the B2B framework, companies can better evaluate the pertinence of storing this data source, especially when customers frequently interact with a focal product or service. Nevertheless, particular emphasis should be placed on developing the necessary expertise to leverage these data optimally and strategically. Based on this concept, the results of this study provide a valuable framework for companies equipped with usage data, while also serving as motivation for those without access to such data.

Last, companies can strengthen their proactive retention strategies in several ways by incorporating usage data. Foremost, by better identifying potential churners through data analysis, firms can emphasize on customers with the highest risk. This approach ensures better allocation of resources by preventing unnecessary disruptions to non-risk customers (Gattermann-Itschert and Thonemann, 2022). In addition, this study provides insights into the underlying motivations for churn, which can be extracted from relevant usage predictors. This deeper understanding enables companies to tailor their retention efforts more precisely by addressing the core causes of churn. For instance, by identifying different expertise levels in the use of a focal product, companies can proactively address knowledge gaps which might lead to a decrease of churn. This approach may also involve personalized communication, loyalty programs, or service improvements that directly address the identified pain points, increasing the likelihood of retaining targeted customers."

2.6.3 Limitations and future research

This study shows the added value of incorporating information on product usage into a CCP. Fine-grained usage data contain valuable information for predicting customer churn, particularly regarding the relevance of timing components. Gaining deeper insights into the conceptual factors that influence the added value of usage data, such as onboarding activities, initial setups, and updates over subscription time, can substantially improve the comprehension of the dynamics associated with churn prediction.

Further investigation of this topic is required to reinforce our findings. Specific insights into usage patterns and their effects on churn prediction may differ in other industries. Future research should use the proposed framework in different fields to validate the significance and impact of usage data in predicting customer churn. Nonetheless, the proposed methodology to structure and assess the added value of this new data source offers managerial insights applicable to a broader range of B2B businesses. Table 2.10 illustrates three B2B settings where the proposed framework can be implemented, along with potential definitions for the three usage components.

Collecting and analyzing usage data is crucial for assessing the value of products and services in different industries. However, resource demands in terms of storage and expertise present significant challenges, especially for smaller companies. Alternatives, such as cloud-based data storage, not only mitigate the burden of physical infrastructure but also offer a scalable and cost-effective choice for data management (Liu et al., 2016). In addition, collaborations with academia, as in this case study, may serve as a strategic path for smaller companies to access

B2B context	Timing	Granularity	Expertise
Marketing agencies	years, semesters, etc.	Project level	Based on requirements (Basic or more advaced)
Telco providers (i.e. Phone)	Days, months, etc.	Call level	Based on used functionalities
Healthcare devices	Days, weeks, etc.	Session level	Based on used functionalities

Table 2.10: Examples of other B2B settings where the framework presented to create usage features can be applied

external expertise and resources that are typically beyond their reach.

Our study successfully demonstrates the predictive power of the proposed models. While acknowledging the value of ongoing refinement, it's important to highlight the predictive power of our models as they outperform a random classifier on all evaluation metrics. Further, the idiosyncratic aspects of a case study, however, might have an impact on the predictive performance of the presented models. For instance, the class distribution can affect the performance of the models, dependent on the specific context of application (Burez and Van den Poel, 2009) as obtaining significant improvements in performance becomes more challenging with lower churn rates. However, in the B2B business context, it is important to denote how even minor improvements in customer churn prediction can translate into significant bottom-line profits due to the high cost of customer acquisition (Rauyruen and Miller, 2007) as demonstrated with the EMPB metric.

Last, future research could delve deeper into more advanced ways to incorporate usage data in CCP models. An interesting avenue for this is the incorporation of usage data into a longitudinal framework instead of cross-sectional one. Deep learning algorithms such as a Long short-term memory (LSTM) network could be explored for this. Such research would extend the stream of literature that investigates innovative ways to combine structured and unstructured data in CCP models (e.g., Benoit and Van Den Poel, 2012; De Caigny et al., 2020).

2.7 Appendix

Appendix A. Notation

	Abbreviation	Explanation
Algorithms (Table 1)	LR	Logistic regression
	SVM	Support vector machine
	RF	Random forest
	DT / C4.5 DT	Decision tree / Specific variation of Decision tree
	NN / NN (MLP)	Neural network / Neural network (Multi layer perceptron)
Metrics (Table 1)	AUC	Area under the ROC curve
	TDL	Top decile lift
	AP	Average precision
	EMPC / EMPB	Expected maximum profit measure for customer churn Expected maximum profit measure for B2B customer churn
Specifications (Table 3)	L1 / L2	Lasso regularization / Ridge regularization
	$\sqrt{F}$	Square root (max. number of features)
	log2(F)	log2(max. number of features)
	min(F,2$\sqrt{F}$)	Minimum (max. number of features, 2$\sqrt{max.numberoffeatures}$))
	min(F,3$\sqrt{F}$)	Minimum (max. number of features, 3$\sqrt{max.numberoffeatures}$
	min(F,4a$\sqrt{F}$})	Minimum (max. number of features, 4$\sqrt{max.numberoffeatures}$

Table 2.11: Table of abbreviations

Appendix B. Example usage features: Usage time

Figure 2.6 illustrates an example of the different combinations of features created from the proposed framework. In the first example, the blocks highlighted represent in this case the average usage time per session for novice actions during the first quarter of the subscription. In the second example, the highlighted blocks represent the total usage time through quarter 4 of the subscription period. The first instance integrates variations across all three defined components—timing, expertise, and granularity. Alternatively, the second example exclusively introduces a timing variation by aggregating usage time in the final quarter.

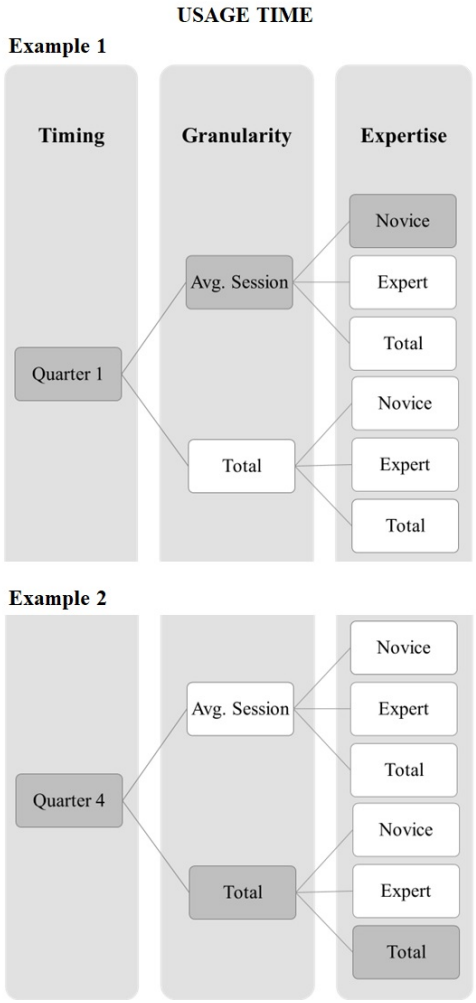


Figure 2.6: Example usage features: Duration of usage

Appendix C. Summary of usage features

The following table presents a summary of the usage features created according to all the possible combinations of the usage components and the usage values, for instance in the case of the last row it is possible to identify the following four features: (1) Average number of expert actions performed per session during the last quarter of the subscription (2) Average usage time per session on expert tasks during the last quarter of the subscription (3) Average document size per session on expert tasks during the last quarter of the subscription (4) Average number of documents per session on expert tasks during the last quarter of the subscription.

Usage Component combinations			**Usage value**			
Timing	**Granularity**	**Expertise**	**Number of actions**	**Usage time**	**Doc size**	**No. Documents**
Total	**Total**	**Total**	✓	✓	✓	✓
Q1	**Total**	**Total**	✓	✓	✓	✓
Q2	**Total**	**Total**	✓	✓	✓	✓
Q3	**Total**	**Total**	✓	✓	✓	✓
Q4	**Total**	**Total**	✓	✓	✓	✓
Total	**Avg. Session**	**Total**	✓	✓	✓	✓
Q1	**Avg. Session**	**Total**	✓	✓	✓	✓
Q2	**Avg. Session**	**Total**	✓	✓	✓	✓
Q3	**Avg. Session**	**Total**	✓	✓	✓	✓
Q4	**Avg. Session**	**Total**	✓	✓	✓	✓
Total	**Total**	**Novice**	✓	✓		
Q1	**Total**	**Novice**	✓	✓		
Q2	**Total**	**Novice**	✓	✓		
Q3	**Total**	**Novice**	✓	✓		
Q4	**Total**	**Novice**	✓	✓		
Total	**Avg. Session**	**Novice**	✓	✓		
Q1	**Avg. Session**	**Novice**	✓	✓		
Q2	**Avg. Session**	**Novice**	✓	✓		
Q3	**Avg. Session**	**Novice**	✓	✓		
Q4	**Avg. Session**	**Novice**	✓	✓		
Total	**Total**	**Expert**	✓	✓	✓	✓
Q1	**Total**	**Expert**	✓	✓	✓	✓
Q2	**Total**	**Expert**	✓	✓	✓	✓
Q3	**Total**	**Expert**	✓	✓	✓	✓
Q4	**Total**	**Expert**	✓	✓	✓	✓
Total	**Avg. Session**	**Expert**	✓	✓	✓	✓
Q1	**Avg. Session**	**Expert**	✓	✓	✓	✓
Q2	**Avg. Session**	**Expert**	✓	✓	✓	✓
Q3	**Avg. Session**	**Expert**	✓	✓	✓	✓
Q4	**Avg. Session**	**Expert**	✓	✓	✓	✓

Table 2.12: Overview of usage features created

Appendix D. Results including all predictors

Usage component	AUC	TDL	EMPB
Timing	**0.615 (0.018)**	**1.677 (0.286)**	**21,148.06 (15,561.69)**
Granularity	0.592 (0.021)	1.420 (0.259)	18,748.36 (18,904.65)
Expertise	0.594 (0.020)	1.420 (0.252)	18,835.42 (17,689.65)
Global features	0.595 (0.021)	1.423 (0.293)	18,539.60 (17,543.77)

Table 2.13: Evaluation metrics per usage components including all features

	AUC		TDL		EMPB	
	F Test	**Adj p-value**	**F Test**	**Adj p-value**	**F Test**	**Adj p-value**
Global features vs. Timing	7.709	**0.029****	5.772	**0.069***	0.185	1.000
Global features vs. Granularity	0.269	1.000	0.043	1.000	0.001	1.000
Global features vs. Expertise	0.029	1.000	0.028	1.000	0.002	1.000

Notes: *The resulting p-values of the F-test are used to compare the performance all usage components in pairwise comparisons* Alpaydin (1998). *Significance levels of 90%, 95%, and 99% are represented by *, **, and ***, respectively, indicating significant differences.*

Table 2.14: Pairwise comparison usage blocks - including all features

	Variables	AUC		TDL		EMPB	
Timing	**Quarter1**	**0.601**	**(0.015)**	**1.694**	**(0.233)**	**20,693.92**	**(17,677.140)**
	Quarter2	0.596	(0.021)	1.669	(0.272)	18,124.92	(15,929.481)
	Quarter3	0.594	(0.023)	1.668	(0.285)	18,230.32	(16,270.915)
	Quarter4	**0.611**	**(0.017)**	**1.749**	**(0.255)**	**19,692.26**	**(15,346.397)**
Granularity	**Session**	**0.599**	**(0.019)**	**1.639**	**(0.203)**	**20,364.860**	**(13,954.590)**
	Total	0.599	(0.018)	1.622	(0.221)	17,576.660	(16,729.380)

Table 2.15: Comparison timing and granularity components - including all features

References

Alpaydin, E. (1998). Combined 5x2cv F Test for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 11:1885–1882.

Anthony, L. F. W., Kanding, B., and Selvan, R. (2020). Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models.

Ascarza, E. and Hardie, B. G. (2013). A joint model of usage and churn in contractual settings. *Marketing Science*, 32(4):570–590.

Barfar, A., Padmanabhan, B., and Hevner, A. (2017). Applying behavioral economics in predictive analytics for B2B churn: Findings from service quality data. *Decision Support Systems*, 101:115–127.

Benoit, D. F. and Van Den Poel, D. (2012). Improving customer retention in financial services using kinship network information. *Expert Systems with Applications*, 39(13):11435–11442.

Blattberg, R. C., Byung-Do, K., and Neslin, S. A. (2008). *Database Marketing*. Springer.

Bloomberg (2023). AMD Gains After Chipmaker Tops Estimates, Makes AI Inroads.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.

Bucklin, R. E. and Gupta, S. (1992). Brand Choice, Purchase Incidence, andSegmentation: An Integrated Modeling Approach. *Journal of Marketing Research*.

Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3 PART 1):4626–4636.

Chen, K., Hu, Y. H., and Hsieh, Y. C. (2015). Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and e-Business Management*, 13(3):475–494.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Z. Y., Fan, Z. P., and Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 223(2):461–472.

Coussement, K. and De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, 66(9):1629–1636.

Coussement, K., Lessmann, S., and Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, 95:27–36.

Coussement, K. and Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1):313–327.

Crone, S. F., Lessmann, S., and Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 173(3):781–800.

De Caigny, A., Coussement, K., De Bock, K. W., and Lessmann, S. (2020). Incorporating textual information in customer churn prediction models based on a convolutional neural network. *International Journal of Forecasting*, 36(4):1563–1578.

De Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K., and Phan, M. (2021). Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. *Industrial Marketing Management*, 99:28–39.

EMBER Climate (2023). Europe Uneven progress towards clean electricity.

EPA (2023). Greenhouse Gases Equivalencies Calculator - Calculations and References.

Erevelles, S., Fukawa, N., and Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2):897–904.

Forrester Consulting (2014). Measuring Customer Health To Drive The Right Conversations. Technical report, Forrester Consulting.

García-Martín, E., Rodrigues, C. F., Riley, G., and Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, 134:75–88.

Gattermann-Itschert, T. and Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*, 107:134–147.

Gordini, N. and Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62:100–107.

Habel, J., Alavi, S., and Heinitz, N. (2023). Effective Implementation of Predictive Sales Analytics. *Journal of Marketing Research*.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1):103–123.

Heberle, H., Meirelles, V. G., da Silva, F. R., Telles, G. P., and Minghim, R. (2015). InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16(1).

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., and Pineau, J. (2020). Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning.

Hochstein, B., Voorhees, C. M., Bolton Pratt, A., Rangarajan, D., Nagel, D., and Mehrotra, V. (2023). Customer Success Management, Customer Health, and Retention in B2B Industries. *International Journal of Research in Marketing*.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Huang, J. and Ling, C. X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. In *IEEE Transactions on knowledge and data engineering*.

Jahromi, A. T., Stakhovych, S., and Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7):1258–1268.

Janssens, B., Bogaert, M., Bagué, A., and Van den Poel, D. (2022). B2Boost: instance-dependent profit-driven modelling of B2B churn. *Annals of Operations Research*.

Lacoste, A., Luccioni, A., Schmidt, V., and Dandres, T. (2019). Quantifying the Carbon Emissions of Machine Learning.

Lakshmanan, A. and Krishnan, H. S. (2011). The Aha! Experience: Insight and Discontinuous Learning in Product Usage. *Journal of Marketing*, 75(6):105–123.

Lee, S. J. and Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data System*, pages 41–46.

Li, D., Chen, X., Becchi, M., and Zong, Z. (2016). Evaluating the energy efficiency of deep convolutional neural networks on CPUs and GPUs. In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 477–484. IEEE.

Lievens, A. and Blažević, V. (2021). A service design perspective on the stakeholder engagement journey during B2B innovation: Challenges and future research agenda. *Industrial Marketing Management*, 95:128–141.

Lilien, G. L. (2016). The B2B Knowledge Gap. *International Journal of Research in Marketing*, 33(3):543–556.

Liu, X., Singh, P. V., and Srinivasan, K. (2016). A structured analysis of unstructured big data by leveraging cloud computing. *Marketing Science*, 35(3):363–388.

Lottick, K., Susai, S., Friedler, S. A., and Wilson, J. P. (2019). Energy Usage Reports: Environmental awareness as part of algorithmic accountability.

Martins, R., Oliveira, T., Thomas, M., and Tomás, S. (2019). Firms' continuance intention on SaaS use – an empirical study. *Information Technology and People*, 32(1):189–216.

Meulemeester, B. and Martens, D. (2023). How sustainable is "common" data science in terms of power consumption? *Sustainable Computing: Informatics and Systems*, 38.

Mora Cortez, R. and Johnston, W. J. (2017). The future of B2B marketing theory: A historical and prospective analysis. *Industrial Marketing Management*, 66:90–102.

Moradi, M. and Dass, M. (2022). Applications of artificial intelligence in B2B marketing: Challenges and future directions. *Industrial Marketing Management*, 107:300–314.

Nevskaya, Y. and Albuquerque, P. (2019). How Should Firms Manage Excessive Product Use? A Continuous-Time Demand Model to Test Reward Schedules, Notifications, and Time Limits. *Source: Journal of Marketing Research*, 56(3):379–400.

Olafsson, S., Li, X., and Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research*, 187(3):1429–1448.

Rafieian, O. and Yoganarasimhan, H. (2022). Variety Effects in Mobile Advertising. *Journal of Marketing Research*, 59(4):718–738.

Rauyruen, P. and Miller, K. E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, 60(1):21–31.

Schweidel, D. A. and Moe, W. W. (2016). Binge Watching and Advertising. *Journal of Marketing*, 80(5):1–19.

Shaw, M. J., Subramaniam, C., Woo Tan, G., and Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31:127–137.

Shi, S. W. and Zhang, J. (2014). Usage experience with decision aids and evolution of online purchase behavior. *Marketing Science*, 33(6):871–882.

Somosi, A., Stiassny, A., Kolos, K., and Warlop, L. (2021). Customer defection due to service elimination and post-elimination customer behavior: An empirical investigation in telecommunications. *International Journal of Research in Marketing*, 38(4):915–934.

Wang, W. Y. C. and Wang, Y. (2020). Analytics in the era of big data: The digital transformations and value creation in industrial marketing. *Industrial Marketing Management*, 86:12–15.

Webster, F. E. and Wind, Y. (1972). *Organizational Buying Behavior*. Prentice-Hall.

Wiersema, F. (2013). The B2B Agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, 42(4):470–488.

Xie, Y., Li, X., Ngai, E. W., and Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3 PART 1):5445–5449.

Zdravevski, E., Lameski, P., Apanowicz, C., and Slezak, D. (2020). From Big Data to business analytics: The case study of churn prediction. *Applied Soft Computing Journal*, 90.

Zhang, J. Z. and Chang, C. W. (2021). Consumer dynamics: theories, methods, and emerging directions. *Journal of the Academy of Marketing Science*, 49(1):166–196.

Zhang, Z., Wang, R., Zheng, W., Lan, S., Liang, D., and Jin, H. (2016). Profit Maximization Analysis Based on Data Mining and the Exponential Retention Model Assumption with Respect to Customer Churn Problems. In *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, pages 1093–1097. Institute of Electrical and Electronics Engineers Inc.

Chapter 3

Dynamic Predictive Modeling of SaaS Trial Conversions: Evidence from a B2B setting

Chapter 3
Dynamic Predictive Modeling of SaaS Trial Conversions: Evidence from a B2B setting

Abstract

This study investigates the extent to which trial-period usage data enhances the prediction of conversion outcomes, thereby supporting more effective, data-driven intervention strategies in SaaS trial-based settings. The analysis draws on 4,566 trial requests collected between April 2022 and December 2024 from a global SaaS provider. We benchmark machine learning models trained on time-invariant usage variables against deep learning models trained on time-varying usage variables. Results of models based on time-varying data significantly improve predictive performance, achieving AUCs above 78% by detecting conversion signals early and refining them as trials progress. To bridge predictive accuracy and behavioral understanding, we analyze usage trajectories and uncover three engagement patterns: Power Users, Steady Adopters, and Disengaging Users. These trajectories show how usage drives conversion probability changes, providing actionable insight into when purchase intent strengthens or weakens. Leveraging these patterns, a Monte Carlo simulation demonstrates that weekly model-guided outreach to high-propensity users generates the largest profit gains. Overall, the findings underscore the importance of incorporating dynamic usage trajectories into predictive targeting to enhance both predictive accuracy and profit potential. By identifying key inflection points where usage patterns signal conversion likelihood, firms can more precisely time interventions, allocate sales resources more efficiently, and ultimately increase conversion rates and profitability.

Keywords: Software-as-a-service, Free-trial conversion, Software usage, Early predictions, B2B

3.1 Introduction

Lead conversion is a critical driver of growth and profitability in Software-as-a-Service (SaaS) markets, where acquiring and retaining paying subscribers determines long-term performance. Managers face the dual challenge of identifying which prospects are most likely to convert and determining when interventions will be most effective. Free trials, a common customer acquisition practice in SaaS settings, provide a structured evaluation period that generates real-time behavioral data valuable for targeting decisions (Datta et al., 2015). Prior research provides examples of how engagement during this period relates to adoption outcomes; for instance, sustained activity has been associated with higher conversion likelihood, whereas excessive feature switching or inactivity have been linked to lower subscription rates (Li, 2022).

Among free-trial designs, time-locked trials are the most common in SaaS (Cheng et al., 2015). By constraining the evaluation window, they introduce urgency and impose a temporal structure on how prospects engage with the product and make adoption decisions. This highlights the importance of understanding engagement as an evolving process rather than a static outcome, making temporality central to trial effectiveness and conversion outcomes. Yet existing research typically measures trial usage through time-invariant variables such as cumulative logins, total active days, or average usage time (Li, 2022; Yoganarasimhan et al., 2023), which obscure within-trial variation and overlook how conversion likelihood changes over time. In contrast, research in areas such as churn prediction and e-commerce conversion shows that time-varying usage dynamics provide stronger predictive power. For instance, (Moe and Fader, 2004) demonstrates that dynamic models built on clickstream sequences outperform static approaches in predicting online conversion.

The absence of similar temporally sensitive measures in SaaS free-trial contexts leaves a critical gap in understanding how usage unfolds within the time-locked window and how this evolution shapes subscription outcomes. This gap constrains both understanding of prospect conversion dynamics and managerial ability to design timely interventions for efficient resource allocation. Addressing this empirical gap requires modeling approaches that capture temporal dynamics and usage complexity while remaining interpretable. However, existing conversion research has largely relied either on cross-sectional classification methods, which condense usage into time-invariant aggregates, or on econometric models that impose restrictive parametric assumptions and require predefined relationships between predictors and outcomes (Meire et al., 2017; Li, 2022). Models trained on cross-sectional data overlook the predictive value of time-varying usage patterns, whereas econometric frameworks, even when applied to time-indexed data, lack the flexibility to capture complex dependencies in high-dimensional settings. We address this limitation by incorporating time-varying usage variables and applying deep learning architectures designed to capture time-varying dependencies, demonstrating that such models generate earlier and progressively more accurate conversion signals.

While methodological advances are necessary to better capture temporal dynamics, interpretability ultimately determines their managerial value. Models that lack transparency face adoption barriers because managers cannot justify resource allocation, design targeted marketing

strategies, or communicate insights effectively to organizational stakeholders (Wu et al., 2024). Effective decision-making therefore requires identifying which prospects are likely to subscribe and understanding the specific usage patterns that signal conversion propensity. Prior research shows that transparent models facilitate adoption by clarifying the behavioral drivers of conversion and demonstrating the consequences of trial design decisions (González-Flores et al., 2025; Yoganarasimhan et al., 2023; Meire et al., 2017). In SaaS contexts, where product usage is central to adoption, interpretability is particularly important for understanding how evolving usage patterns relate to subscription outcomes and for guiding timely, targeted interventions throughout the trial period. This study addresses this gap by complementing conversion prediction models with an interpretable clustering approach that identifies three prospect profiles, each one associated with different usage trajectories and conversion probabilities. Further managerial relevance is demonstrated through a profit simulation that contrasts dynamic interventions during the trial with single end-of-trial interventions, highlighting the financial implications of timely conversion strategies.

This research contributes to the SaaS literature through an empirical study that examines how time-varying usage variables enhance conversion prediction during free trials. The results show that incorporating time-indexed variables significantly improve predictive power compared to prospect characteristics or time-invariant usage variables. We demonstrate that time-varying variables yield earlier, more accurate, and more actionable signals of conversion likelihood. The integration of predictive modeling with usage profiling underscores the relevance of accuracy and interpretability, revealing how distinct engagement patterns during the trial correspond to different conversion probabilities. Profit simulations indicate that weekly, model-guided interventions generate higher net returns than single end-of-trial approaches or random weekly selection, highlighting the economic value of adaptive prospect conversion efforts. Taken together, these findings support the conceptualization of trial engagement as a dynamic process, providing actionable guidance for managers in designing intervention strategies that are timely, targeted, and resource-efficient.

The remainder of this paper is organized as follows. Section 3.2 presents a review of literature relevant to SaaS Lead conversion, usage data, and modeling approaches. Section 3.3 outlines the empirical setting and data sources. Section 3.4 describes the predictive framework developed to train and test the different models. Section 3.5 presents the results of the predictive models, comparing algorithms and data sources. Section 3.6 provides interpretable insights from the predictive models and describes the usage patterns found. Section 3.7 details the simulation framework used to assess the profit impact of alternative targeting strategies. Finally, Section 3.8 concludes the paper with the relevant managerial implications and overall findings.

3.2 Literature Review

3.2.1 SaaS Lead conversion prediction

SaaS represents a fundamental shift from traditional software licensing models, replacing one-time purchases with recurring subscriptions to software applications (Loukis et al., 2019). This transition fundamentally alters the customer acquisition process, as the lower barriers to trial adoption and the recurring revenue model create both opportunities and challenges for lead conversion. Unlike traditional enterprise software sales, where conversion decisions are typically binary and final, SaaS models allow for extended evaluation through free trials, freemium offerings, and gradual feature adoption, creating prolonged decision processes and multiple conversion touchpoints (Mehra and Saha, 2018).

Research, however, has offered limited guidance on how providers manage acquisition under these conditions. Studies of SaaS adoption have largely taken a demand-side perspective, emphasizing organizational, technological, and environmental drivers of adoption (Oliveira et al., 2019; Seethamraju, 2015). Provider-side work, in turn, has concentrated on pricing and operations (Li and Kumar, 2022), while offering comparatively little insight into customer acquisition and, in particular, lead conversion. Yet lead conversion is fundamental to growth in subscription-based businesses, since free trials, demonstrations, and digital campaigns generate large prospect pools that must be systematically prioritized.

This operational challenge has led to the adoption of lead scoring methodologies, in which firms use predictive models to estimate the likelihood of conversion and allocate scarce sales resources to the most promising opportunities (González-Flores et al., 2025). Early rule-based approaches, often grounded in managerial intuition, have been criticized as arbitrary and prone to bias, frequently leading to inefficient allocation of effort (Järvinen and Taiminen, 2016). Predictive lead scoring seeks to overcome these limitations by leveraging historical data and advanced analytics to identify attributes most strongly associated with conversion, thereby standardizing evaluation processes and mitigating bias. These models typically integrate multiple data sources that capture different dimensions of the prospect and the organization: firmographic information, such as company size, industry, and geography (Meire et al., 2017; D'Haen et al., 2013), and online behavioral data, which capture dynamic engagement signals from digital interactions such as website activity, email responses, and social media engagement (Järvinen and Taiminen, 2016; Li, 2022).

Existing research provides limited insights into how lead conversion prediction can be implemented in SaaS, particularly with respect to the development and evaluation of predictive models for guiding acquisition decisions. This methodological gap is especially relevant in B2B SaaS settings, where complex acquisition processes, extended sales cycles, and high contract values increase the stakes of predictive accuracy (Järvinen and Taiminen, 2016). This study addresses this gap by developing analytical insights into lead conversion prediction in B2B SaaS, comparing diverse predictive approaches and assessing their relative performance in estimating conversion likelihood.

3.2.2 Trial Usage Data in Conversion Prediction

Within this landscape, usage data from free trials has emerged as a particularly distinctive source of insight, offering direct behavioral signals of interest and engagement that complement firmographic and digital traces.

Free trials are not only a marketing tool but also a learning environment in which consumers update their beliefs and valuations through direct product experience (Cheng et al., 2015). Usage during the trial helps consumers compare their initial expectations against realized performance, thereby shaping their true valuation of the product. In this sense, consumer learning is central: innovator users, such as those in beta programs, refine their preferences and knowledge through hands-on engagement, and prospective buyers improve their beliefs about product quality only after actual use (Lin et al., 2024). Importantly, usage entails associated costs, as consumers must invest effort and time to acquire familiarity with software functionalities (Cheng et al., 2015). These usage costs constitute an integral component of the evaluation process and subsequently influence conversion likelihood.

Beyond shaping consumer beliefs and valuations, usage patterns also provide measurable behavioral metrics that have been incorporated into conversion prediction models. Usage data capture the intensity, diversity, and continuity of interactions, offering direct behavioral signals of engagement that shape conversion outcomes. Research shows that higher levels of usage reinforce consumer learning, reduce quality uncertainty, and strengthen commitment, thereby increasing retention, repeat purchasing, and the persistence of engagement over time (Bolton and Lemon, 1999; Lemon et al., 2002; Ascarza and Hardie, 2013; Sriram et al., 2015; Datta et al., 2015; Sanchez Ramirez et al., 2024). Previous research has incorporated trial usage data into predictive models of lead conversion, particularly in the context of free trials for digital services. These studies consistently emphasize that usage metrics evidence consumer engagement and learning during the trial period, shaping the likelihood of subscription or retention.

In the SaaS context, Li (2022) explore trial usage into two behavioral dimensions, frequency, measured by the number of software launches, and variety, measured by the number of distinct applications explored. Their findings show that higher cumulative frequency is positively associated with conversion, while greater variety is negatively associated, suggesting that excessive exploration may dilute focus and reduce subscription likelihood. In a complementary perspective, Yoganarasimhan et al. (2023) examine trial design influences on conversion through measures such as functionalities used, active days, and dormancy length, demonstrating that longer trials increase conversion through consumer learning while extended inactivity proves detrimental. These findings establish usage intensity as a fundamental predictor of conversion outcomes, with behavioral engagement during trials serving as a critical signal of purchase intent.

However, most studies operationalize trial usage through aggregated or low-frequency measures that overlook how engagement unfolds within the trial period. Yoganarasimhan et al. (2023) employ time-invariant usage variables that summarize engagement across entire periods, whereas Li (2022) use a dynamic hazard model with decaying stock variables of frequency and variety. While this model captures cumulative engagement, it does not incorporate intra-trial dynamics that reflect how usage patterns evolve during evaluation. These approaches thus provide

useful summaries but fail to capture the temporal progression of prospect engagement. By contrast, research in customer retention demonstrates that incorporating timing and progression of usage—including recency, inactivity intervals, trends, and sequence effects, substantially improves prediction compared to time-invariant aggregates (Moe and Fader, 2004; Ascarza and Hardie, 2013). Extending this temporal perspective to lead conversion would allow prediction models to account for evolving engagement during trials, offering insight into prospect learning and conversion likelihood. Yet current approaches remain restricted to time-invariant variables rather than time-varying sequences, leaving intra-trial dynamics underexploited in SaaS conversion prediction. The current study addresses this gap by developing a predictive framework that leverages daily usage data, capturing activity, feature usage, dormancy, variety, and user count, to improve the representation of trial engagement and its role in lead conversion prediction.

3.2.3 Temporality in Lead Prediction Models

Understanding how engagement likelihood unfolds over time is essential for accurate lead conversion prediction, yet the temporal dimension remains underexplored in SaaS contexts. Li (2022) advance the literature by modeling cumulative advertising and usage effects through exponentially decaying stock variables in a dynamic hazard framework. While this approach captures carryover effects, it imposes restrictive functional forms that may not reflect the nonlinear trajectories of digital customer journeys, which are becoming increasingly prevalent in settings such as B2B. Yoganarasimhan et al. (2023) provide evidence on how trial design shapes conversion outcomes, but their framework does not specify how conversion probabilities can be dynamically updated as prospects advance through the trial, limiting its operational relevance. More recently, González-Flores et al. (2025) demonstrate the value of machine learning models for lead scoring, yet their cross-sectional approach overlooks the temporal evolution of engagement and the need for continuous recalibration as buyer behaviors and market conditions change.

Taken together, prior research reveals a persistent gap, as existing approaches either aggregate trial behavior into static measures or impose restrictive parametric structures, thereby limiting their ability to capture the evolving nature of engagement during evaluation. The present study addresses this limitation by incorporating dynamic predictive variables that reflect how usage patterns unfold over time, which in turn enables conversion probabilities to be updated as new information becomes available. This study compares traditional machine learning techniques with deep learning architectures designed for the integration of time-varying data, applying a rolling-horizon evaluation that mirrors the ongoing progression of trials. The study advances methodological understanding of temporal prediction while also providing practical guidance for prospect prioritization in dynamic prospect conversion efforts.

3.2.4 Interpretability in Lead Prediction Models

Prior research on lead conversion highlights interpretability as central to linking predictive performance with business-relevant insights. Meire et al. (2017) show how variable importance identify concrete predictors of acquisition success while explaining differences in model performance,

making complex models actionable for managers. Similarly, González-Flores et al. (2025) employ feature importance and diagnostics to isolate key attributes such as lead source, classification, and state, which reflect customer behavior and inform sales strategies. These approaches establish interpretability as essential for operational adoption, enabling managers to understand not only which prospects are likely to convert but also why certain characteristics drive conversion. Interpretability has also been advanced through studies that uncover mechanisms explaining conversion outcomes. Yoganarasimhan et al. (2023) apply causal inference and segmentation analyses to trial length, identifying two countervailing mechanisms—consumer learning and dormancy—that jointly determine subscription outcomes and vary across user segments. Similarly, Li (2022) demonstrates how advertising touchpoints and trial usage interact to shape subscription likelihood. In addition, to illustrate managerial implications, Li (2022) implements simulations that vary the timing of one additional display impression or email among consumers who have already been exposed. These simulations highlight windows in which incremental marketing effort can raise conversion, but they do not consider prospects with no prior exposures, nor do they incorporate costs or profitability.

While these contributions extend interpretability beyond static predictors and reveal important behavioral mechanisms, a gap remains. Prior studies clarify why and under what conditions customers convert, but they provide limited guidance on how evolving engagement trajectories, especially in the absence of marketing contact, can be transformed into operational decision rules. In practice, managers must decide not only who to target but also when to intervene and how much behavioral information is needed before acting. Our study addresses this gap by demonstrating that time-varying usage data yield earlier and more reliable signals of conversion likelihood than end-of-trial measures. Through the integration of predictive models with these usage profiles and profit simulations of intervention strategies, we show that weekly, model-guided interventions outperform both random targeting and end-of-trial campaigns. This extends interpretability from marginal exposure effects to dynamic engagement processes, providing managers with evidence-based frameworks for optimizing timing, targeting precision, and resource allocation in prospect conversion strategies.

3.3 Empirical setting and Data

We partner with a B2B SaaS provider specializing in pre-press editing software, operating as an independent firm within a Fortune 500 company. The main software product, designed to prepare PDF files for printing, is offered under a yearly subscription-based pricing model. Before purchase, prospective customers can access a 30-day free trial by submitting a request form available on the company's website. This trial provides unrestricted access to the full functionality of the software.

The data used in this study consist of all 4,566 trial requests collected between April 2022 and December 2024. Prospect data include geographic region, intended product use, and operating system, providing basic information about each prospect before any product interaction. In contrast, usage data capture granular engagement patterns based on time-stamped interaction logs.

These two sources jointly enable the construction of variables at multiple temporal resolutions, supporting different granularity configurations. Appendix 1 includes an extract from the usage logs containing a few sample entries.

This empirical context is well-suited for exploring dynamic prediction strategies, in which the model incorporates time-varying data to update predictions continuously as user behavior unfolds. Implementing such an approach benefits from a dataset that integrates temporally detailed usage logs with prospect characteristics. These data enable the construction of time-dependent variables that capture changes in usage engagement throughout the trial period and support multiple temporal aggregation scales, such as time-invariant and time-varying variables, allowing evaluation of conversion prediction across varying degrees of granularity. This temporal resolution underpins a modeling framework capable of adapting predictions as new usage information becomes available. The following sections detail how these data are transformed into predictive variables through targeted feature engineering.

3.3.1 Prospect Data

In this setting, a prospect firm can request a free trial from the focal company by completing an online registration form. The form requires the prospect to provide contact information, details about its organization, the intended use case for the software, and preferences regarding follow-up communication. Upon submission of a trial request, the prospect organization receives a unique activation key to install and access the software for a 30-day trial period. The information provided during registration remains fixed throughout the trial and forms the basis for a subset of the prospect variables used in the analysis. Only relevant variables for predictive modeling were included, while all unique identifiers such as company name, telephone number, and other information that could directly identify the organization were removed.

The prospect variables obtained from the registration form are limited to geographic region and stated usage intention. The latter was classified into three mutually exclusive categories: Checking, Editing, and Fixing PDF files. In addition to these registration-based variables, the operating system of the device used to run the free trial was captured from the usage logs. Although not collected through the original request form, the operating system is considered a prospect variable because the trial software can only be installed and activated on a single machine per trial. All three prospect variables, geographic region, usage intention, and operating system, are categorical and were dummy-encoded for inclusion in the predictive modeling process.

The prospect data provides valuable insights into the description of prospects who request free trials. As shown in Figure 3.1a, most trial activations originate from Europe (45.4%) and North America (26.6%), followed by Asia (14.1%), with smaller proportions from South America, Oceania, and Africa. Finally, Figure 3.1b presents the distribution of stated usage intentions, segmented into three key groups: Checking (32.9%), Editing (36.5%), and Fixing PDF files (30.6%), evidencing the heterogeneity of customer needs during the trial period. Appendix 2 presents a summary of all the prospect variables and their values.

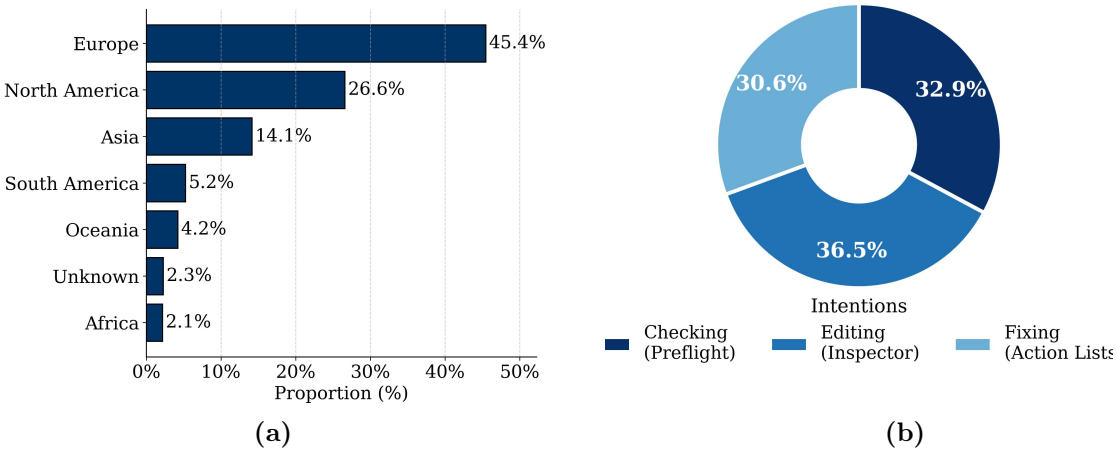


Figure 3.1: (a) Distribution per Region. (b) Distribution of software-usage intentions.

3.3.2 Usage Data

Usage data are derived from log files, which record all activities and events performed by a user. During the trial period, each account is activated with a unique activation key, which enables the tracking of usage data and its association with a specific user. The usage data are organized into several distinct categories, including application-level information, document-related actions performed on the processed file, hardware data describing the system environment, license details, execution logs documenting the use of specific software features, and preflight reports generated during PDF validation processes. Together, these logs provide a detailed and time-stamped record of user behavior during the trial period.

These data form the basis for the creation of variables, producing a comprehensive set of time-invariant and daily time-varying variables. To ensure accurate analysis and comparison of usage data at the daily and trial levels, five primary usage metrics, based on prior literature (Yoganarasimhan et al., 2023; Li, 2022), characterize prospect behavior during the trial period: (1) Usage activity, indicating when the software was used; (2) dormancy, measured as the number of consecutive days without usage; (3) action variety, capturing the number of distinct action types performed; (4) number of active users; and (5) frequency of usage per action type. Logged user actions are categorized into six functional groups, each representing a distinct function of the software: *Document* includes operations such as opening, saving, or exporting files; *Preflight* involves validation of the PDF document prior to printing; *Edit* refers to actions aimed at automating instructions; *Inspector* comprises functions that enable users to select and examine individual elements of the file; *Tools* allow for manual modifications used during the review process; and *Other* includes less frequent actions not classified under the previous categories. Figure 3.2 illustrates examples of these categories, showing actions associated with Preflight to validate PDF files, Inspector to examine and modify elements, and Tools for direct content adjustments.

To provide a broader view of the user interaction workflow, Figure 3.3 illustrates the variable

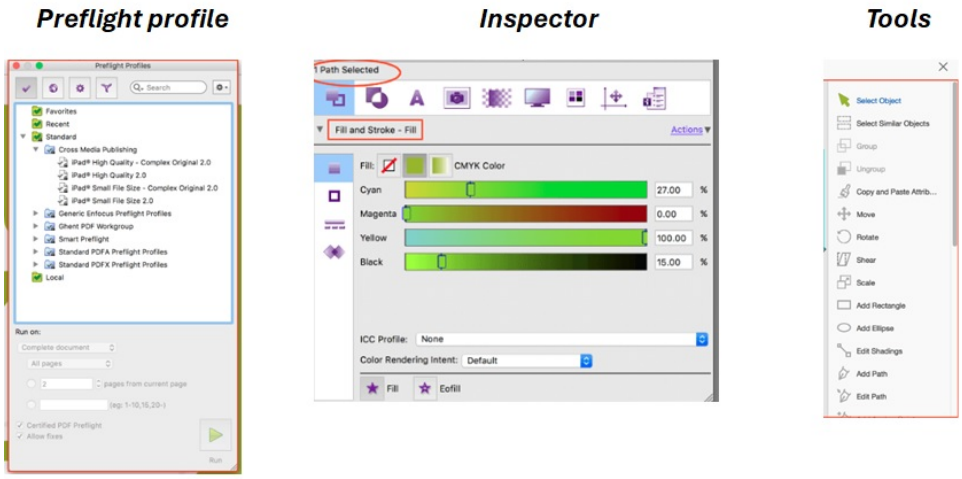


Figure 3.2: Examples action categories in the software

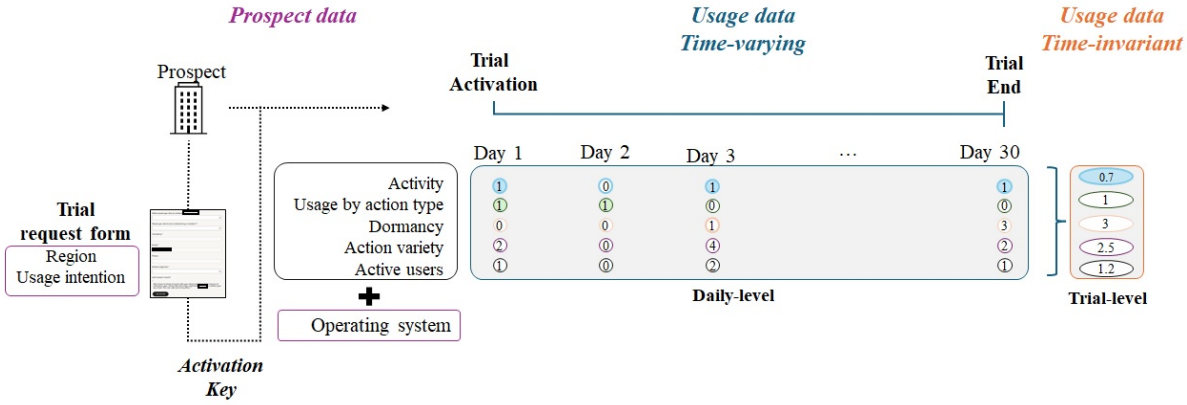


Figure 3.3: Overview of Trial Process and variable creation

collection process starting from the moment a prospect submits a trial request. The figure also depicts how the five usage metrics are aggregated at the daily and trial levels, providing the foundation for assessing the added value of dynamic usage data. Time-varying aggregation captures user behavior on a per-day basis, preserving the temporal dynamics of trial usage. This level of granularity allows for modeling the evolution of usage behaviors and conducting time-based analyses of user engagement with the software. In contrast, time-invariant aggregation summarizes user behavior over the entire 30-day period, producing an aggregated set of variables for each trial observation. More precisely, at the daily level, the activity metric is represented as a binary indicator of use, which at the trial level is summarized as the ratio of total active days. Dormancy is also first computed daily, as the number of consecutive inactive days, and then aggregated at the trial level as the consecutive inactivity period by the end of the subscription. In the same way, action variety, frequency of usage by action type, and active users are measured on a daily basis and subsequently aggregated to the trial level by taking the maximum value observed during the trial period.

Table 3.1 presents descriptive statistics, including the mean, minimum, maximum, and standard deviation. The descriptive statistics show a pronounced heterogeneity in user engagement

Feature	Mean (μ)	St. Dev. (σ)	Max	Min
Active rate	0.2	0.2	1.0	0.0
Days Since Last Use	14.2	11.6	30.0	0.0
Variety	5.0	1.5	7.0	1.0
User Count	1.1	0.3	7.0	1.0
Action Categories				
Execute	20.8	113.5	6133.0	0.0
Document	30.2	89.7	1831.0	0.0
Preflight Report	48.5	297.3	16870.0	0.0
Edit	8.3	28.9	690.0	0.0
Inspector	59.5	210.8	7493.0	0.0
Tools	26.9	208.2	10166.0	0.0
Others	87.1	297.9	8282.0	0.0
Active Days	6.4	6.4	30.0	0.0

Table 3.1: Descriptive statistics usage variables

during the trial period. Metrics such as the active rate (μ= 0.2, σ= 0.2) and days since last use (μ= 0.2, σ= 11.6) reflect varying intensity and recency of software interaction across users. The wide variation in action counts, particularly in the Execute, Preflight Report, and Inspector categories, suggests that while some prospects engage extensively with these functions, others make minimal use of them or do not engage at all. Conversely, low mean values in categories like *Edit* suggest these functions are used sporadically or by a limited subset of users.

Figure 3.4 examines whether users predominantly engage with the specific software feature they initially intended to use. It presents daily usage trajectories over the 30-day trial for users segmented by their stated feature intention, Preflight Report, Inspector Tool, or Action List Tool. The results indicate that users display a moderate alignment between stated intentions and actual engagement, with the degree of correspondence varying across feature types. Among prospects with a stated intention to use Preflight (a), usage of the Preflight Report is initially highest, indicating some alignment with intention; however, usage declines over time and remains comparable to or slightly lower than Inspector Tool usage during the remainder of the trial. In contrast, those indicating an intention to use the Inspector Tool (b) consistently exhibit higher usage of the Inspector Tool throughout the 30-day period, reflecting the strongest correspondence between intention and behavior among the three groups. For users who reported an intention to use the Action List Tool (c), the observed pattern diverges: initial usage is relatively high but declines rapidly, remaining below both Preflight and Inspector usage throughout the trial period. This trajectory suggests limited sustained engagement with the feature of interest.

Overall, the analysis shows that engagement with the stated feature of interest frequently diminishes over time. However, prospects indicating an intention to use the Inspector Tool maintain steady use of Inspector and emerge as the group with the highest cumulative engagement with this feature among the three intention categories. These patterns may reflect users' evolving assessments of tool utility, usability, or task relevance as they explore the software during the trial. Such behavioral dynamics underscore the importance of complementing self-reported intention data with longitudinal usage data to more accurately capture engagement patterns and inform

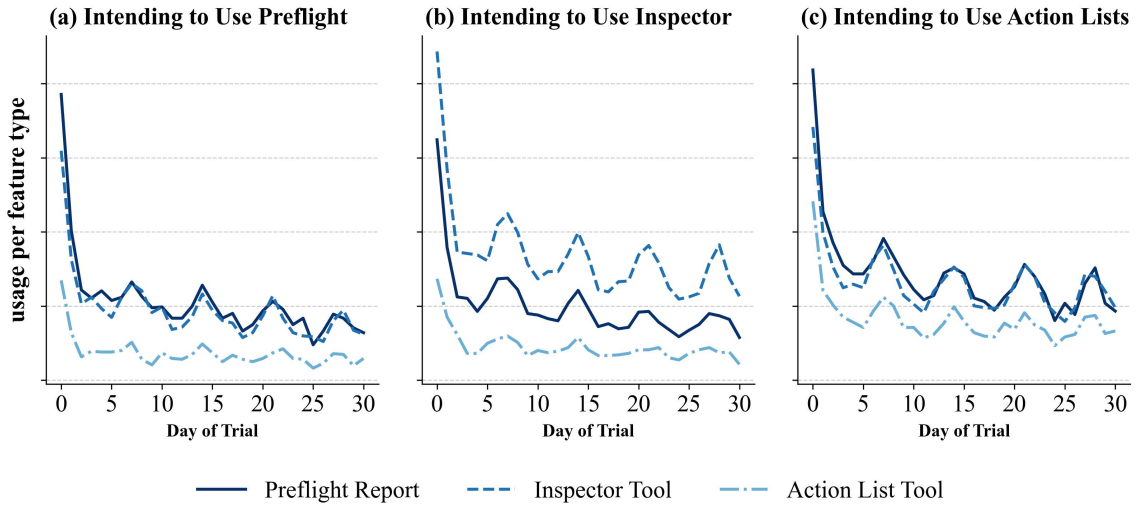


Figure 3.4: Daily usage by stated usage intention.

predictive modeling. Appendix 2 presents a summary of all the usage variables and their detail levels.

3.3.3 Conversion and descriptives

Upon activation of the free trial, prospects may choose to upgrade to a paid subscription at any time during the trial period or after its completion. To rigorously identify conversions attributable to the trial experience, we implement a tracking window of 90 days from trial start. This three-month observation horizon is consistent with the focal firm's business practices for evaluating free-to-paid transitions and is further supported by empirical evidence. As illustrated in Figure 3.5, which presents the cumulative proportion of eventual converters by days elapsed since trial start, approximately 72.5% of all conversions occur within this 90-day period. This finding substantiates the use of the three-month cutoff while ensuring alignment with established industry practice.

Conversions occurring beyond the 90-day window are likely influenced by factors unrelated to the initial trial experience, such as subsequent marketing activities, customer interactions, or targeted conversion campaigns. To limit the influence of such factors on the measurement of trial conversions, the attribution window is set at 90 days. Within this period, the binary outcome variable takes the value of one if a prospect subscribes and zero otherwise.

Figure 3.6a indicates meaningful differences in stated usage intentions between converters and non-converters. A greater proportion of converters (42%) selected Inspector as their intended use than non-converters (34%), whereas Preflight was more frequently selected by non-converters (36%) than by converters (26%). The Action List intention category remains relatively similar across both groups. This suggests that users interested by the Inspector feature are more likely to convert, highlighting usage intention as a potential predictor of conversion. These differences in intention are mirrored in actual usage patterns over time. As presented in figure 3.6b, across all action categories, ranging from general activity levels to specific interactions such

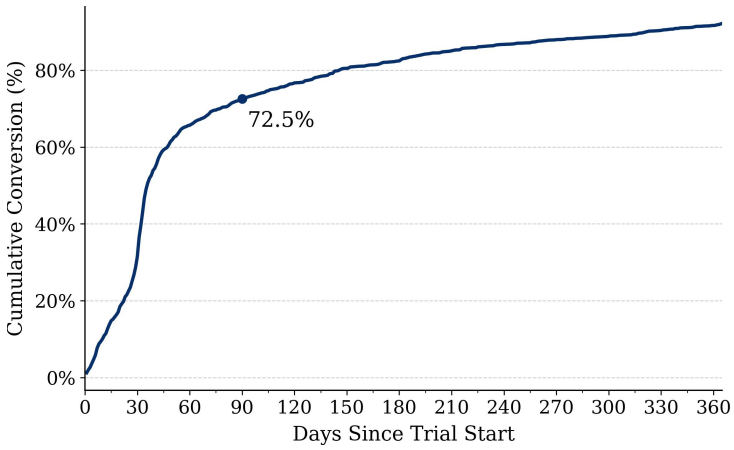


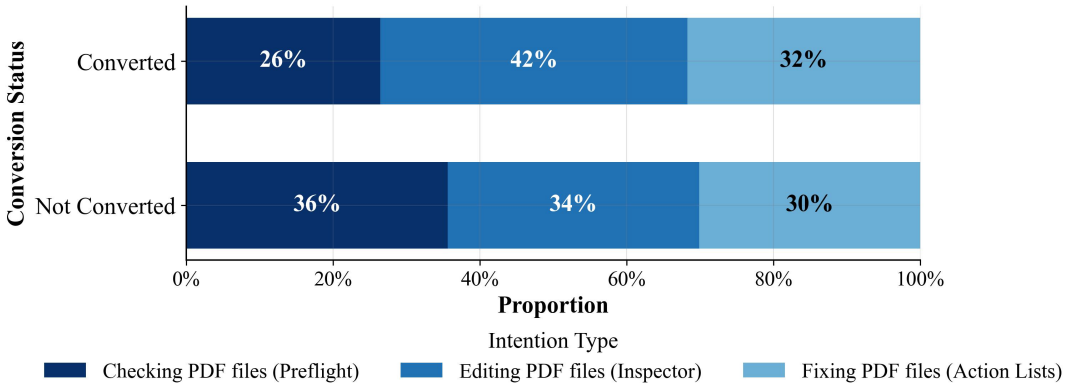
Figure 3.5: Cumulative distribution of conversion over time

Feature	Mean (μ)		St. Dev. (σ)		Max		Min		Trend (30d)	
	Converted	Not Converted	Converted	Not Converted	Converted	Not Converted	Converted	Not Converted	Converted	Not Converted
Active Rate	0.3	0.2	0.2	0.2	1.0	1.0	0.0	0.0		
Days Since Last Use	8.4	16.8	9.7	11.4	30.0	30.0	0.0	0.0		
Variety	5.5	4.8	1.4	1.5	7.0	7.0	1.0	1.0		
User Count	1.1	1.0	0.4	0.3	7.0	7.0	1.0	1.0		
Active Days	10.0	4.7	7.1	5.3	30.0	29.0	1.0	0.0		
Execute	31.6	15.9	183.3	58.7	6,133.0	1,510.0	0.0	0.0		
Document	51.9	20.5	125.3	65.4	1,831.0	1,195.0	0.0	0.0		
Preflight Report	81.5	33.6	499.2	122.5	16,870.0	3,249.0	0.0	0.0		
Edit	11.4	6.8	37.4	23.9	690.0	532.0	0.0	0.0		
Inspector	103.2	39.8	309.0	142.0	7,493.0	2,896.0	0.0	0.0		
Tools	46.0	18.3	218.3	203.0	5,528.0	10,166.0	0.0	0.0		
Others	147.5	59.8	382.1	246.1	7,444.0	8,282.0	2.0	0.0		

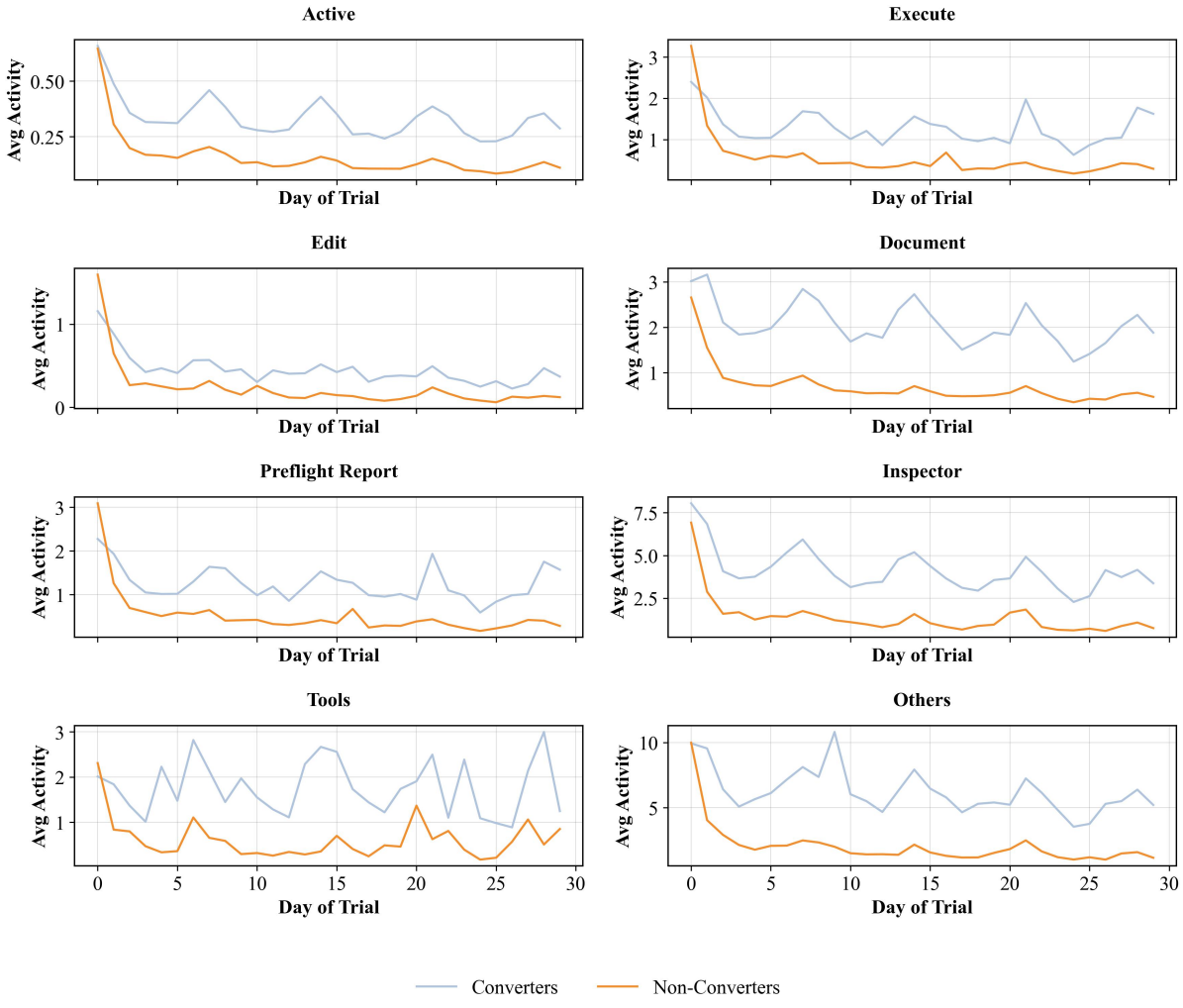
Table 3.2: Usage metrics vs. conversion

as executing commands, editing, document handling, and generating preflight reports, converted users consistently demonstrate higher average activity throughout the 30-day trial and display systematic weekly seasonality. For instance, converters exhibit pronounced early engagement, followed by sustained activity, whereas non-converters display lower initial activity and a more rapid decline. The Inspector and Others categories show disparities in intensity, suggesting that frequent use of inspector or miscellaneous tools may correlate with conversion likelihood.

Descriptive statistics from Table 3.2 reinforce these observations. Converted users exhibit higher means across nearly all variables, including active rate, variety of actions, and number of active days. The days since last use for converters are notably lower on average, implying more recent engagement towards the trial's end. Furthermore, trend plots reveal that converters maintain steadier or less declining usage patterns compared to the more variable and generally decreasing activity of non-converters. Together, these findings emphasize the importance of both the nature of usage intentions and the intensity and persistence of software interaction in differentiating converters from non-converters. Integrating these detailed temporal and behavioral patterns into predictive models is likely to significantly enhance their performance in estimating prospect conversion.



(a) Usage intention vs. conversion



(b) Usage activity per actions over time

Figure 3.6: Overview of conversion metrics

3.4 Predictive framework and Models

To investigate how different forms of data contribute to trial conversion prediction, we develop a structured predictive framework comprising five distinct model specifications that vary in their inclusion of prospect, time-invariant, and time-varying usage variables. Section 3.4.1 outlines this predictive framework and the rationale for each model configuration, enabling a systematic analysis of the incremental predictive value offered by different variable types and their combinations. Section 3.4.2 then describes the different types of algorithms used to implement these models, including both machine learning models for prospect data and time-invariant usage data and deep learning approaches tailored to capture temporal patterns in time-varying usage data.

3.4.1 Predictive Framework

As illustrated in Table 3.3, we specify five distinct model configurations to systematically evaluate the individual and combined predictive value of prospect and usage-based variables. Our analytical approach begins with a benchmark model, designated as *Model A*. It incorporates only prospect variables to represent who the prospect is prior to any product interaction. This model quantifies the baseline predictive accuracy attainable using only basic information.

	Prospect Data	**Usage**	
		Time-invariant	**Time-varying**
Model A	✓		
Model B		✓	
Model C			✓
Model D	✓	✓	
Model E	✓		✓

Table 3.3: Overview predictive of models.

To isolate the contribution of usage data, we estimate two specifications at distinct temporal resolutions. *Model B* uses time-invariant aggregates to assess how overall engagement relates to conversion. *Model C* uses daily variables to capture usage dynamics and tests whether these dynamics add predictive value. Comparing the predictive performance of *Models B* and *C* with the prospect-only baseline (*Model A*) quantifies the incremental value of usage information, while contrasting *Model B* and *Model C* reveals how temporal resolution of usage variables affects conversion prediction.

To further explore the predictive value of usage data, we extend our analysis by examining whether prospect and usage variables offer complementary information when integrated. The final two models are specified to assess the joint contribution of these two data sources. *Model D* augments the baseline prospect variables with the time-invariant usage aggregates introduced

in *Model B*. This setting allows us to evaluate whether time-invariant usage variables provide predictive value extending beyond the information contained in prospect profile. Building on this, *Model E* combines the prospect variables from *Model A* with the time-varying usage variables from *Model C*. This specification evaluates how prospect variables contribute to predictive accuracy when combined with temporal patterns in usage behavior.

3.4.2 Predictive models

To develop predictive models that identify whether a company initiating a free software trial subsequently converts to a paid subscription, we frame this task as a binary classification problem. The input data contain a combination of prospect, time-invariant usage, and time-varying usage variables recorded throughout the 30-day trial. To evaluate the predictive value of dynamic usage data, we design the experiment in two distinct phases. The first phase applies classical machine learning algorithms to prospect and time-invariant usage variables, while the second phase implements deep learning models designed to exploit temporal patterns in time-varying usage data.

Machine learning

The first modeling strategy relies on prospect and time-invariant usage variables in a cross-sectional manner. Classical machine learning methods are well-suited for this setting, as they effectively handle tabular data and capture structured patterns within time-invariant inputs. Three algorithms previously tested in trial conversion literature (Yoganarasimhan et al., 2023), were implemented in this phase: logistic regression, Random forest, and XGBoost (Hastie et al., 2009).

Logistic regression (LR) is a widely used classification technique that models the probability of a binary outcome as a function of input variables through a logistic function. It assumes a linear relationship between the predictors and the log-odds of the target variable. While relatively simple in formulation, LR remains a competitive algorithm across various applications, offering a strong baseline for predictive modeling. Its simplicity and interpretability make it particularly valuable for understanding the marginal effects of individual variables (Chou et al., 2022), compared to more complex methods. In the context of digital subscription and purchase modeling, it has been effectively employed to examine the influence of user engagement and platform touchpoints on conversion outcomes (Li, 2022).

Random forest (RF) is an ensemble learning method based on decision trees, which introduces two layers of randomness: bootstrapping the training data and randomly selecting subsets of variables at each split (Breiman, 2001). This design improves model generalization and mitigates overfitting, making it well-suited for complex classification tasks involving mixed data types and non-linear relationships. RF has demonstrated strong performance in customer behavior modeling, particularly in contexts requiring the integration of behavioral and demographic variables (Reiners et al., 2025; Somanchi et al., 2025).

XGBoost (XGB) is a gradient boosting algorithm that builds an ensemble of weak learners,

typically shallow decision trees in a sequential manner. At each step, the model focuses on correcting the errors made by its predecessors, optimizing an objective function that balances prediction accuracy and model complexity (Chen and Guestrin, 2016). This iterative refinement allows XGB to capture subtle patterns and complex variable interactions, often yielding superior performance in classification tasks. Boosting algorithms have been widely adopted in marketing applications, such as consumer behavior prediction, purchase intent and subscription conversion (Rafieian and Yoganarasimhan, 2021; Martínez et al., 2020; Baumann et al., 2018). In the context of trial-based conversion, Yoganarasimhan et al. (2023) emphasizes the potential of advanced machine learning methods like XGB to inform the design and evaluation of optimal trial strategies.

Deep learning

While traditional machine learning algorithms are effective at capturing patterns in prospect tabular data, they often fall short when applied to longitudinal or time-varying information. In such contexts, more advanced deep learning models are better suited to capture the temporal dependencies and complex patterns inherent to time-varying information. Given the nature of the data in this case study, two deep learning algorithms well-suited for prediction tasks with time-varying data were implemented: a one-dimensional Convolutional Neural Network (1D-CNN) and a Long Short-Term Memory network (LSTM). These models are designed to leverage the temporal structure of the input data and are expected to provide deeper insights into how day-to-day engagement influences conversion outcomes.

1D-CNNs offer an alternative approach for processing time-varying data by applying convolutional filters across time steps to detect local patterns in feature activity. By extracting spatially localized temporal variables through hierarchical feature maps, 1D-CNNs leverage a feedforward architecture that offers greater computational efficiency (Kiranyaz et al., 2015). This makes them a competitive choice for modeling time-varying usage interactions, especially when long-term memory is less critical than pattern recognition. Their effectiveness in predicting customer behavior has been demonstrated across domains such as e-commerce and finance (Chaudhuri et al., 2021; Khan and Niu, 2021).

LSTM is a type of recurrent neural network (RNN) specifically designed to model time-varying data by learning long-range dependencies across time steps(Graves et al., 2013). It addresses the limitations of traditional RNNs, such as vanishing and exploding gradients, through a gating mechanism that regulates the flow of information via input, forget, and output gates (Hochreiter and Schmidhuber, 1997). This architecture enables LSTM networks to retain relevant information over extended sequences, making them well-suited for applications involving user behavior over time. In the context of trial-based conversion prediction, LSTM can effectively model the temporal dynamics of time-varying usage patterns, capturing how the timing and consistency of interactions influence subscription decisions. Previous studies in purchase prediction and behavioral modeling have highlighted the advantage of LSTM in extracting temporal dependencies from clickstream or usage data (Sakar et al., 2019; Wang et al., 2024).

3.4.3 Training and Cross-validation

To ensure that all predictive models were both robust and generalizable, a nested cross-validation strategy was employed uniformly across modeling approaches. The outer loop used a 10-fold cross-validation procedure to assess model performance on unseen data, while the inner loop applied a 5×2 cross-validation scheme for hyperparameter tuning. This nested design separates model evaluation from parameter optimization, thereby reducing the risk of overfitting and avoiding biased performance estimates. Such a validation strategy is particularly well-suited for datasets of moderate size, where single random splits may produce unreliable or non-representative assessments of predictive accuracy (Hastie et al., 2009).

Hyperparameter tuning was performed within the inner loop using parameter ranges informed by prior empirical studies (see Appendix 3). This process was applied consistently across all models, including both traditional machine learning algorithms and deep learning architectures. Machine learning models were trained using the different variable combinations of time-invariant data (Models A, B, and D), with evaluation based solely on the outer loop to ensure unbiased performance metrics. For the deep learning models, input data were formatted as three-dimensional tensors of shape (number of prospects, time steps, variables), where each tensor slice represented the daily sequence of activity variables for a given prospect. To account for variability in sequence lengths—arising from inconsistent user activity or missing daily observations—sequences were zero-padded to a fixed length of 30 days. Padding was temporally aligned to the trial start date, preserving the chronological structure essential for effective temporal modeling. In model E, where both dynamic and prospect inputs were available, prospect variables required to be integrated into both LSTM and 1D-CNN architectures. For the LSTM model, prospect variables were concatenated with the output of the sequential encoding layer. Specifically, the final hidden state of the LSTM was combined with prospect variables before being passed to the prediction layer. For the 1D-CNN model, time-varying variables were first processed with a 1D convolution and flattening operation to extract temporal patterns, and then concatenated with prospect variables in a fully connected layer prior to prediction. This design enabled both model types to jointly capture temporal dynamics and time-invariant variables, enhancing predictive performance without compromising sequence learning.

To assess the performance of the model in predicting trial conversion, the primary evaluation metric used was the Area Under the Receiver Operating Characteristic Curve (AUC). AUC is a widely accepted metric for binary classification tasks, as it measures the ability of the model distinguishing between converted and non-converted users across all possible classification thresholds, it provides a threshold-independent assessment of discriminative power.

3.5 Predictive Lead Conversion Results

Table 3.4 reports the validation AUC scores associated with each variable configuration. Additionally, Appendix 4 presents the fine-grained 5x2-CV results, including the mean and standard deviation of the AUCs.

Model	Logistic Regression		Random Forest		XGBoost	
	Mean AUC	St. dev	Mean AUC	St. dev	Mean AUC	St. dev
(A) Prospect Data	62.89%	0.03	**63.30%**	0.03	63.10%	0.03
(D) Prospect + Time-invariant usage	77.59%	0.02	77.51%	0.02	**78.21%**	0.02
(E) Prospect + Time-varying usage	77.30%	0.02	76.77%	0.02	**77.87%**	0.02

Table 3.4: Comparison of test AUC across subsets of variables.

In line with the experimental framework outlined above, the results are divided into four sections. First, *Predictive Value of Usage Information* asseses whether incorporating behavioral usage data, beyond prospect variables, improves predictive performance. Second, *Added Value of Time-Varying Usage Data in conversion prediction* examines the impact of usage data granularity by comparing models that use time-invariant with those incorporating time-varying variables. Third, *Comparative Performance of Modeling Techniques* evaluates how model performance varies across different algorithmic approaches, including traditional machine learning and deep learning techniques. Finally, *Temporal Prediction Gains with LSTM* investigates whether models designed to capture time-varying patterns, specifically the LSTM architecture enables early prediction by leveraging daily time-varying usage data.

3.5.1 Predictive Value of Usage Information

This analysis examines the incremental predictive value of behavioral usage data in comparison to prospect variables. As prospect variables are inherently cross-sectional and lack temporal structure, only traditional machine learning methods are applied in this section. Table 3.4 shows that incorporating behavioral usage data substantially improves predictive accuracy. Models trained exclusively on prospect variables (Model A) consistently produce the lowest AUC scores, 62.89% for Logistic Regression, 63.30% for Random Forest, and 63.10% for XGBoost, establishing a baseline for evaluating the added value of usage information.

Incorporating usage information, whether at the trial or daily level (Models D and E), substantially improves predictive performance across all machine learning models. For instance, the inclusion of time-invariant usage variables increases AUC values to 77.59% for Logistic Regression, 77.51% for Random Forest, and 78.21% for XGBoost. This consistent improvement suggests that user engagement during the trial period is a strong predictor of conversion offering greater predictive value than prospect variables alone.

The improvement in predictive performance from incorporating usage data is statistically significant. Appendix 5 reports Wilcoxon signed-rank test results (Wilcoxon, 1945), providing pairwise comparisons between models using only prospect variables and those combining time-invariant and time-varying usage variables. The tests confirm that, across all models, the inclusion of usage data leads to a significant increase in predictive accuracy. These findings offer robust evidence that behavioral usage data significantly improves conversion prediction in trial-based settings.

3.5.2 Added Value of Time-Varying Usage Data in conversion prediction

Having established the value of incorporating behavioral usage data, we next investigate whether the granularity of that information, specifically the distinction between aggregated time-invariant metrics and time-varying usage variables yields further performance improvements. To address this question, both traditional machine learning and deep learning models are included in the comparison. While traditional algorithms allow to incorporate prospect and time-invariant variables, deep learning architectures offer the potential to better capture the added value of time-varying usage data, particularly due to their ability to model temporal dependencies when such structure is present in the input.

Model	Logistic Regression		Random Forest		XGBoost		1D-CNN		LSTM	
	Mean AUC	St. dev	Mean AUC	St. dev	Mean AUC	St. dev	Mean AUC	St. dev	Mean AUC	St. dev
(B) Time-invariant usage	74.99%	0.03	**75.76%**	0.03	75.55%	0.03	–	–	–	–
(C) Time-varying usage	75.28%	0.03	75.76%	0.02	**75.59%**	0.03	74.30%	0.03	**76.73%**	0.03
(E) Prospect + Time-varying usage	77.30%	0.02	76.77%	0.02	**77.87%**	0.02	74.72%	0.02	**78.30%**	0.02

Table 3.5: Comparison of test AUC across subsets of variables usage granularity.

As presented in Table 3.5, for traditional machine learning models, time-varying and time-invariant usage variables yield nearly identical predictive performance. Both approaches achieve a maximum AUC of 75.76% with Random Forest, while XGBoost performs similarly, reaching 75.59% with time-varying inputs and 75.55% with time-invariant inputs. However, the advantage of daily usage data becomes pronounced when deep learning architectures are employed. The LSTM model, designed to capture temporal dependencies, attains an AUC of 76.73% using only daily-level usage data. This performance surpasses that of all traditional machine learning models using either aggregated or fine-grained inputs. Furthermore, when prospect variables are combined with time-varying usage data (Model E), the LSTM model achieves the highest AUC observed in the study (78.30%), underscoring the complementary value of prospect information, fine-grained behavioral data, and temporal modeling. Although descriptive results suggest modest performance gains from using time-varying variables instead of time-invariant aggregates, Wilcoxon-Holm tests (Appendix 6) indicate that these differences are not statistically significant for traditional machine learning models. This implies that increasing data granularity alone does not necessarily enhance predictive performance unless the model architecture can effectively capture the temporal structure within the data. In contrast, LSTM-based models, which are explicitly designed to model time-varying dependencies, significantly outperform all traditional machine learning alternatives when applied to time-varying usage data, with adjusted p-values below 0.05 in each comparison. These findings highlight the importance of pairing temporally detailed data with architectures capable of modeling behavioral dynamics, reinforcing the value of deep learning approaches in settings where user engagement evolves throughout the trial period.

3.5.3 Comparative Performance of Modeling Techniques

The comparative analysis of algorithms unpacks the respective strengths and limitations of different modeling approaches in extracting predictive signals from usage data. Among traditional machine learning models, XGBoost consistently delivers the highest AUCs, reaching up to 78.21% (Table 3.4 Model D) when both prospect and time-invariant usage variables are incorporated, and 77.87% (Table 3.4 Model E) with prospect and time-varying usage variables. Random Forest and Logistic Regression also benefit from richer variable sets, although their performance remains slightly lower.

Deep learning models, particularly the LSTM architecture, demonstrate strong capabilities in capturing time-varying patterns in user behavior. When trained on prospect and time-varying usage data, the LSTM model achieves an AUC of 78.3% (Table 3.5, Model E), consistently outperforming all alternative modeling approaches. While the 1D-CNN also leverages rich input data, achieving an AUC of 74.72% under the same setting, its performance remains below that of the LSTM, potentially due to the latter's greater capacity to capture temporal dynamics.

These results suggest that while advanced machine learning models such as XGBoost can approach the performance of deep learning models when provided with comprehensive, structured input, the LSTM architecture is particularly well-suited to leveraging temporal patterns in user behavior. This capacity to model behavioral dynamics enhances its effectiveness in operational contexts that require early and accurate identification of high-potential users for timely, targeted interventions.

3.5.4 Temporal Prediction with LSTM

To evaluate the value of temporally structured prediction, we implement a dynamic modeling framework that generates weekly predictions based on cumulative daily usage data observed up to the corresponding week. This approach reflects a real-time decision environment in which firms aim to identify likely converters as early as possible, updating predictions as additional usage data accumulate over the course of the trial.

As established in the preceding analysis, the LSTM model outperforms all other approaches when trained on time-varying usage data combined with prospect variables. Its superior performance underscores its capacity to capture temporal patterns that unfold over the trial period. This finding motivates the use of LSTM in the dynamic prediction framework, where the goal is to generate weekly predictions based on cumulative daily usage data. By preserving the temporal order of user activity and updating predictions as new data become available, LSTM enables a more responsive modeling approach that aligns with the evolving nature of user engagement during the trial.

The results in Figure 3.7 reveal a clear upward trajectory in predictive accuracy as the trial progresses. After just 7 days, the LSTM model achieves an AUC of 71.5%, already outperforming models based solely on prospect variables. Its performance continues to improve with additional usage data, reaching 74.9% by day 14, 75.6% by day 21, and peaking at 78.3% by the end of the 30-day trial, surpassing all benchmark machine learning models trained on prospect and

time-invariant usage data. These findings highlight the value of cumulative, temporally structured inputs in enabling increasingly accurate predictions and facilitating the early identification of likely converters. The performance gains further underscore the practical relevance of sequence-aware modeling in dynamic usage environments.

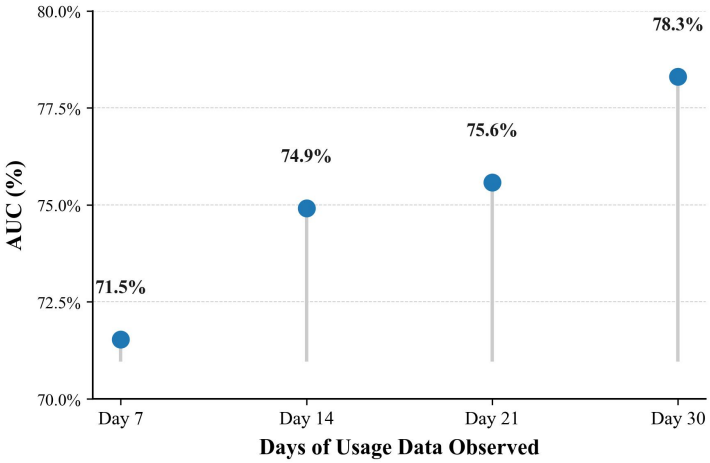


Figure 3.7: Dynamic AUC performance of the LSTM model based on cumulative daily usage data observed at weekly intervals.

3.6 Interpretable insights

To examine the temporal usage patterns underlying conversion prediction outcomes, we implement a behavioral segmentation analysis. While the LSTM models demonstrate that temporally structured data enable early and accurate identification of likely converters, they offer limited interpretability with respect to the behavioral dynamics associated with conversion prediction. To address this gap, we implement a two-stage clustering approach adapted from Wang et al. (2024), in order to investigate temporal variation in user engagement during the trial period.

The segmentation framework captures user heterogeneity arising from temporal patterns of usage. In the first stage, k-means clustering is applied to the time series of each usage variable to identify representative behavioral patterns over the 30-day trial period. These cluster assignments capture recurring temporal structures at the variable level. In the second stage, k-modes clustering is performed on the set of variable-level cluster labels, resulting in user-level assignments to higher-order behavioral segments. This two-stage approach produces interpretable segments based on dynamic engagement profiles while avoiding the interpretive complexity associated with clustering raw time-series data directly.

The resulting clusters reveal three distinct patterns of engagement behavior, as visualized in the time-series plots in Figure 3.8a and summarized in Table 3.6. Each cluster is characterized by a unique temporal behavior that reflects different modes of interaction with the software over the trial period. The first cluster, that we denote as *Steady Adopters*, includes users who maintain moderate but consistent activity levels across all behavioral dimensions. Their use of variables such as document editing, execution, and inspector access remains relatively stable, and their

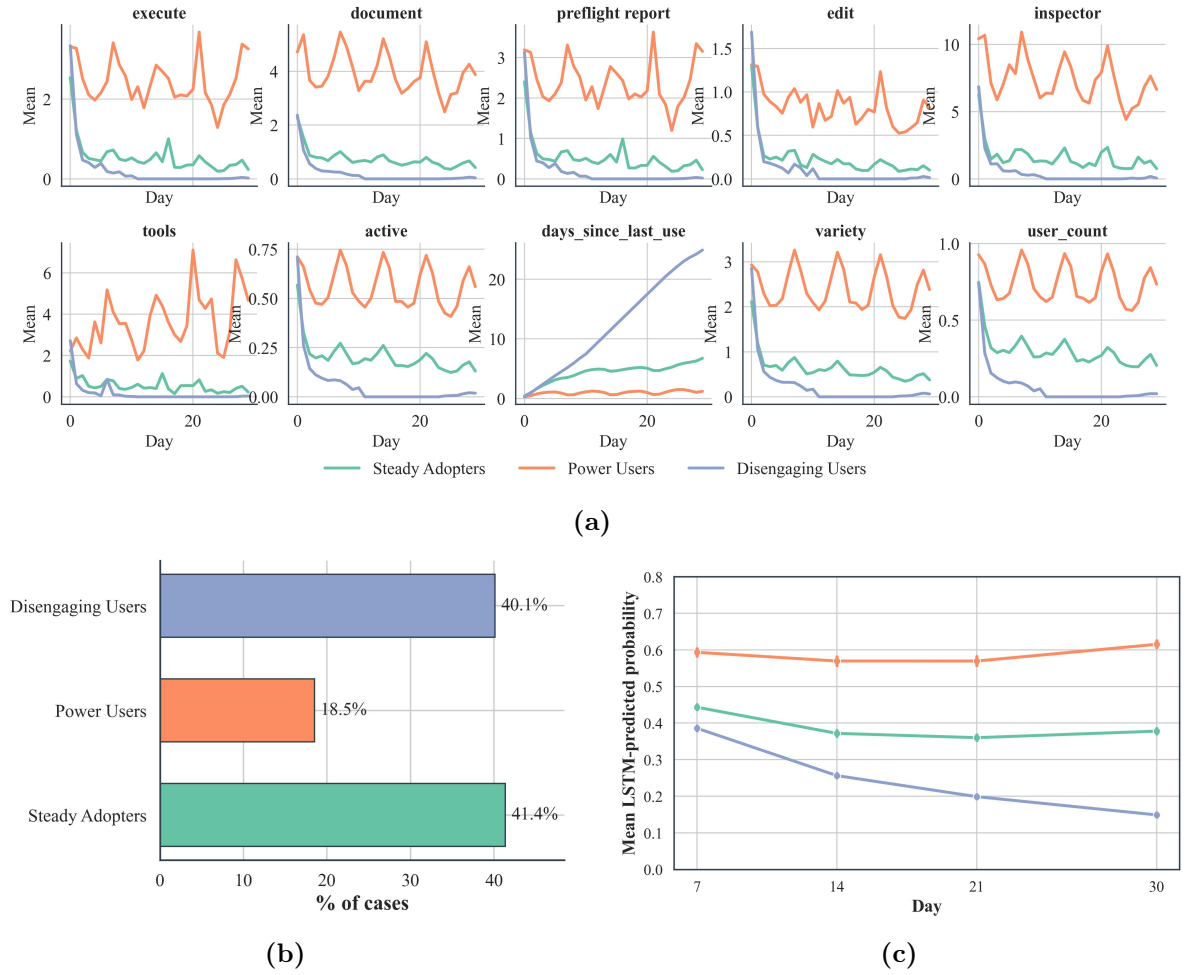


Figure 3.8: Clustering results and model-predicted probabilities. (a) Temporal centroids by cluster; (b) Cluster distribution; (c) LSTM-predicted probabilities by cluster.

Features		All	Steady Adopters	Power Users	Disengaging Users
Time-varying usage variables	Execute	0.8	0.6	2.5	0.2
	document	1.1	0.7	4	0.2
	Preflight report	0.7	0.5	2.4	0.2
	Edit	0.3	0.2	0.8	0.1
	Inspector	2.2	1.6	7.4	0.5
	Tools	1	0.5	3.7	0.2
	Active days	0.2	0.2	0.6	0.1
	Days since last use	6.9	4.3	1	12.3
	Variety	0.8	0.6	2.4	0.2
	User count	0.3	0.3	0.7	0.1
Prospect variables	os_ARM-based_max	0.2	0.2	0.2	0.2
	OS (ARM-based)	0	0	0	0
	OS (Intel-based)	0.2	0.3	0.2	0.2
	OS (Intel-based Macintosh)	0	0	0	0
	OS (Unknown)	0	0	0	0
	Usage intention Checking (Preflight)	0.2	0.2	0.1	0.2
	Usage intention Editing (Inspector)	0.2	0.2	0.2	0.2
	Usage intention Fixing (Action Lists)	0.2	0.2	0.2	0.2
Conversion rate		31.1%	36.2%	58.2%	13.3%

Table 3.6: Feature descriptives per each k-mode cluster.

"days since last use" metric stays low throughout the trial period. This group reflects a pattern of regular but not intensive engagement, serving as a baseline for comparison with the other segments.

In contrast, users in the cluster *Power Users* exhibit sustained, high-frequency engagement across all usage dimensions. Their behavioral patterns are marked by repeated peaks in activity and a consistently low "days since last use" metric, indicating uninterrupted interaction with the platform throughout the trial. This group reflects intensive and continuous usage behavior.

Finally, the segment of *Disengaging Users* follows a distinct engagement pattern marked by an early spike in activity, followed by a steady and sustained decline across all behavioral variables. Their "days since last use" metric increases rapidly, indicating early and persistent disengagement. Their early and lasting inactivity identifies them as a high-risk group for churn. One plausible interpretation is that these users may engage with the tool to complete a single, short-term task, without an intention for ongoing use. Alternatively, early disengagement may stem from unmet expectations or dissatisfaction with the product experience. Distinguishing between these underlying causes has important implications for targeting retention strategies and designing differentiated onboarding or support mechanisms.

The segmentation provides a structured framework for interpreting the output of the LSTM model. As shown in Figure 3.8c, predicted conversion probabilities differ across clusters and over the trial period. *Power Users*, characterized by consistently high engagement across multiple features, receive the highest predicted probabilities throughout. *Steady Adopters*, with moderate and stable usage, are assigned intermediate probabilities throughout the trial. *Disengaging Users*, whose usage engagement declines rapidly after initial use, show a pronounced decrease in predicted probabilities after the first week. These patterns suggest that variation in usage behavior over time aligns with the predictions of the conversion model. While the model does not rely on the segment labels, the clustering results offer a useful context for interpreting how temporal engagement relates to predicted conversion likelihoods.

As summarized in Table 3.6, the consistency between predicted conversion probabilities and actual outcomes reinforces the validity of the behavioral segmentation. Conversion rates vary substantially across the identified clusters: while the overall rate in the sample is 31.1%, it increases to 58.2% among *Power Users* and 36.2% among *Steady Adopters*, but falls to 13.3% among *Disengaging Users*. To assess whether these behavioral profiles correspond to other underlying user characteristics, we compare the distribution of prospect variables such as operating system (OS) and stated usage intention across clusters. The results reveal minimal variation, suggesting that the primary source of heterogeneity lies in temporal usage behavior rather than demographic or contextual factors. These findings support the view that prospect variables alone are insufficient for identifying meaningful patterns in conversion behavior. In contrast, dynamic behavioral data provide more granular and actionable insights into user heterogeneity in trial-based settings.

The heterogeneity observed in usage behavior, highlights the importance of treating user engagement as a dynamic process. This perspective has practical implications for the timing of interventions, as it suggests that response strategies should be informed by evolving behavioral

patterns rather than prospect variables. Accordingly, dynamic usage data can be leveraged not only to classify users into behavioral segments but also to guide the timing of marketing interventions. In the following section, we examine how predictive models can be integrated with temporal usage insights to support more effective, data-driven intervention strategies in trial-based settings.

3.7 Dynamic Intervention Simulation

The cumulative impact of advertising exposure across multiple touchpoints is well documented in the marketing literature. A robust body of empirical work demonstrates that aggregate advertising impressions delivered through display, email, or search channels have a positive and significant impact on consumer conversion rates (Naik and Raman, 2003; Zantedeschi et al., 2017; Danaher et al., 2020). Extending this work, Li (2022) provides direct empirical evidence of these effects within free-trial settings in the SaaS industry. Analyzing detailed user-level data and marketing touchpoints, the study demonstrated that cumulative marketing exposure across digital channels consistently increased the likelihood that users would transition from trial to paid subscriptions. In this section, we simulate dynamic intervention strategies to explore how predictive targeting can be adapted over time, providing leads on implementation by identifying how marketing efforts can be optimized across prospects and decision points.

Motivated by these findings, we develop a simulation framework to evaluate the business impact of dynamic, data-driven intervention strategies in trial-based settings. Building on the dynamic conversion probabilities estimated in section 3.5.4, we construct a Monte Carlo simulation to quantify the operational and financial implications of deploying model-guided marketing interventions. This simulation translates predictive outputs into decisions about *when* to intervene during a 30-day trial. It compares four strategies that differ in targeting method, either random or model-based, and in intervention timing, implemented as a single contact at the end of the trial or as weekly contacts distributed throughout the trial period.

3.7.1 Intervention policies

To evaluate the operational and financial implications of dynamic targeting, the simulation compares four distinct intervention strategies. Each policy is constrained to reach approximately 25% of the customer population, ensuring that all strategies operate under equivalent budgetary conditions. This constraint reflects practical limitations where marketing or customer success resources are typically restricted and cannot be deployed universally. Similar constraints are commonly assumed in the targeting literature, for instance, Ascarza (2018) evaluates intervention policies under fixed outreach capacities to reflect real-world budgetary constraints in retention contexts.

The four intervention strategies summarized in Figure 3.9, differ along two primary dimensions: the approach to user selection, which includes either random assignment or model-based targeting; and the timing of marketing contacts, defined as either a single outreach at the end of the trial

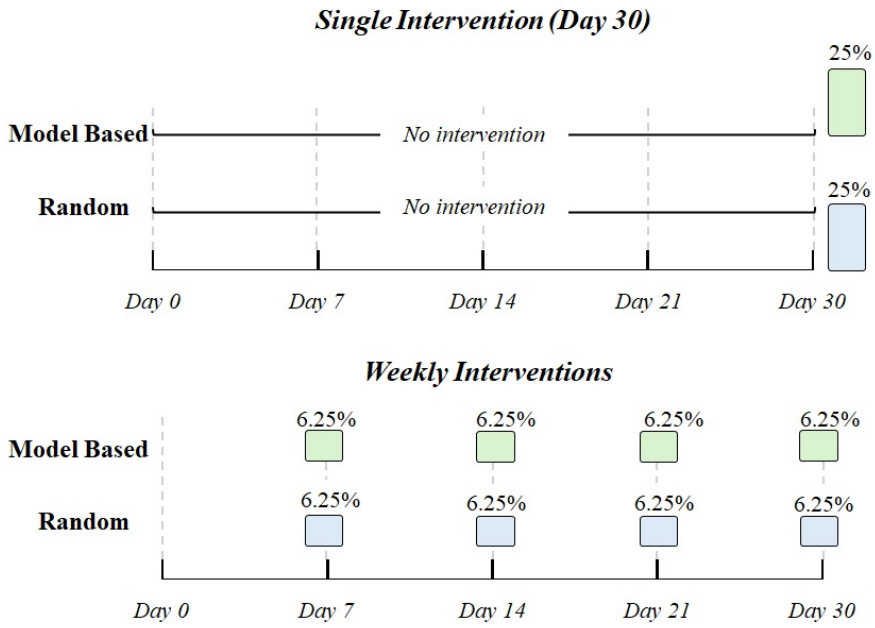


Figure 3.9: Simulation scenarios.

period or multiple staggered contacts distributed throughout the trial. Across all scenarios, each prospect is eligible to receive at most one marketing contact. In the *Random (Day 30)* strategy, 25% of users are selected uniformly at random to receive a one-time marketing contact on the final day of the trial. The *Random (4 Weeks)* strategy distributes this contact volume evenly over four weeks by selecting 6.25% of users at random each week, excluding individuals already contacted in prior weeks and allowing for weekly interventions while maintaining budget parity.

In contrast, the *Model-Based (Day 30)* strategy uses predicted conversion probabilities at day 30 to target a specific quartile of the population, delivering a one-time contact at that point. The *Model-Based (4 Weeks)* strategy extends this approach through dynamic segmentation: each week, updated predictions are used to select the 6.25% of prospects for contact, excluding individuals already contacted in prior weeks. This process results in a total contact volume of 25% by the end of the trial, while allowing intervention decisions to adapt over time based on evolving user behavior.

3.7.2 Impact in conversion

Building on prior work, Li (2022) shows that targeted messaging based on usage patterns can significantly enhance free trial conversion. Reported conversion uplifts from related email studies span a relatively modest range: triggered browse-abandonment emails increase purchase probability by about 0.2 percentage points (Goic et al., 2021), while personalized subject lines raise lead conversion rates by roughly 0.12 percentage points (Sahni et al., 2018). Guided by this literature, we adopt $u = 0.07$ as our baseline simulation parameter. This value lies within the range of previously documented effects, represents a conservative magnitude, and aligns with the finding from Li (2022) about the effectiveness of usage-targeted communications. To assess robustness, we additionally report sensitivity analyses with alternative u values in Appendix 7.

To capture the incremental effect of marketing interventions, we define the treated conversion probability $\tilde{p}_i$ as a function of the baseline conversion probability estimated at the policy-defined contact time and whether the prospect received a marketing contact during the trial period. Specifically, for each targeted prospect i, the post-intervention conversion probability is given by $\tilde{p}_i = \min\{1,\ \hat{p}_{i,30} + u \cdot k\}$. Where $\hat{p}_{i,t}$ denotes the estimated baseline probability at the contact time t determined by the intervention policy, u represents the additive uplift per contact, and $k \in \{0, 1\}$ indicates whether a marketing contact was received.

We conduct a simulation to illustrate the application of an uplift-based marketing intervention. Specifically, we adjust the baseline conversion probabilities $\hat{p}_{i,t}$ defined above and add a fixed uplift u for prospects in the treatment group. Because conversions are binary outcomes, we generate simulated conversions as $y_i \sim \text{Bernoulli}(\tilde{p}_i)$. This approach preserves the discrete nature of the outcome, introduces random variation at the prospect level, and allows repeated runs to assess how different targeting strategies would perform.

3.7.3 Profit assessment

To evaluate financial performance in all intervention strategies, we compute total profit Π as the difference between revenue generated from successful conversions and the cumulative cost of marketing contacts. For each simulation run, profit is defined as:

$$\Pi = \sum_{i=1}^{n} \left(R\,\tilde{y}_i - C\,m_i\right), \tag{3.1}$$

where $\tilde{y}_i \in \{0, 1\}$ is the simulated conversion outcome for user i, and m_i denotes the number of marketing contacts received by user i over the trial period. The revenue parameter is set to $R = 400$, based on the yearly subscription cost for the case of study. The cost per contact is set at $C = 15$, reflecting operational expenses per outreach and consistent with estimates in a B2B SaaS context reported by Janssens et al. (2022). This formulation provides a net profit measure that reflects both the financial return from conversions and the implementation cost of the corresponding intervention strategy.

Each strategy is evaluated over 1,000 Monte Carlo replications to capture variation due to stochastic outcomes and to obtain a robust distribution of the profit metric. Statistical comparisons between strategies are performed using paired two-sample t-tests, which assess whether observed differences in mean profit are statistically significant under repeated sampling.

3.7.4 Simulation Results

To determine which users to target under the outreach capacity constraint, we divide the sample into quartiles ordered by predicted conversion probability. Model-based interventions are then evaluated separately within each quartile. The profit-maximizing quartile Q4 is designated as the intervention group, with performance details reported in Appendix 8. All subsequent simulations fix the intervention group to this quartile to ensure consistent and interpretable comparisons across policies.

Figure 3.10a reports the average percentage change in profit relative to the no-intervention baseline across all intervention strategies under an uplift parameter of $u = 0.07$. The results show clear performance differences by targeting and timing: model-based strategies consistently outperform random allocation, and the dynamic weekly policy (M4W) achieves the highest profit increase. Figure 3.10b illustrates the underlying mechanisms by decomposing contacted prospects into three categories based on the simulation outcomes: already converters, non-responders, and new converters. Because the outreach budget is fixed, all policies reach the same number of prospects, yet model-based strategies allocate contacts more effectively, generating substantially more incremental conversions. M4W, for example, yields on average 302 incremental conversions, compared with 272 for MD30 and fewer than 240 for both random strategies. These results demonstrate that predictive targeting reduces wasted contacts on prospects who would convert without intervention or fail to respond, thereby concentrating resources on prospects with higher marginal responsiveness. The additional gains of M4W over MD30 further underscore the value of temporal adaptation: by updating predictions weekly, the policy identifies evolving signals of conversion and redirects outreach to the most promising users. Overall, the decomposition shows not only why model-based strategies deliver superior profits but also how dynamic reallocation transforms a fixed outreach budget into substantially greater incremental returns.

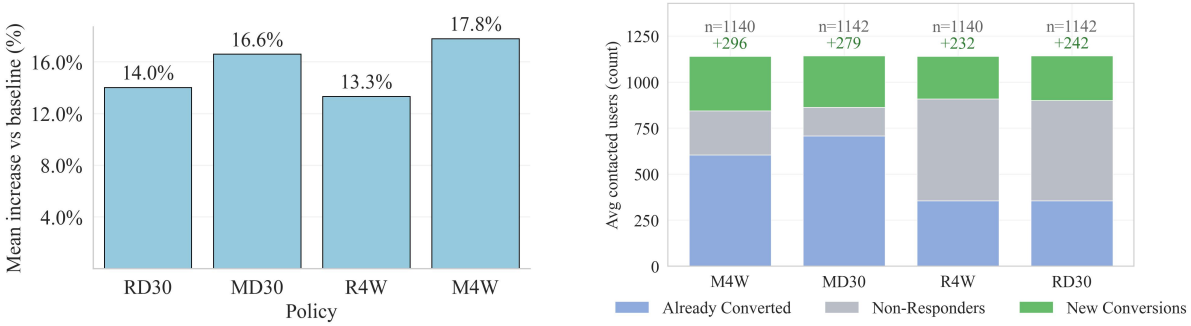
Under a fixed outreach budget of 25% and an assumed uplift of $u = 0.07$, all intervention policies generate positive profit relative to no intervention, yet their effectiveness differs depending on both targeting and timing. The results demonstrate that predictive targeting is the primary driver of value, shifting from random to model-based selection at Day 30 increases profit by 2.7 percentage points. Timing further enhances effectiveness, but its value depends on the quality of targeting. While distributing contacts over time yields only a marginal gain under random selection (0.4 percentage points), it produces an additional improvement when combined with model-based predictions (2.1 percentage points). Consequently, the model-based weekly strategy (M4W) emerges as the dominant policy, achieving an 18.2% profit increase, outperforming both its single-shot counterpart (MD30, 16.1%) and the best random strategy (R4W, 13.8%). Because contact costs are held constant across strategies, these differences reflect more efficient allocation of limited outreach capacity, with dynamic, data-driven targeting ensuring that marketing efforts are directed toward prospects with the highest incremental likelihood of conversion. This finding underscores the operational value of coupling predictive models with sequential interventions, highlighting how information updates over the trial period can be systematically leveraged to maximize financial returns.

Table 3.7 reports paired t-tests comparing the performance of the different policies. Additional sensitivity analyses under alternative uplift values of 5% and 9% are provided in Appendix 9. Across all comparisons, the differences are statistically significant at the 0.1% level, with mean profit gaps ranging from roughly \$2,300 (RD30 vs. R4W) to more than \$27,600 (RD30 vs. M4W). These results confirm that both predictive targeting and pacing contribute significantly to financial performance, with the model-based weekly strategy (M4W) consistently dominating all alternatives. This finding underscores the economic value of dynamically updating predictions and allocating scarce outreach capacity to the prospects with the highest incremental potential

Scenario 1	Scenario 2	Mean Diff	p-value
RD30	MD30	-15770	0.000 ***
RD30	R4W	-2267.2	0.000 ***
RD30	M4W	-27658.4	0.000 ***
MD30	R4W	13502.8	0.000 ***
MD30	M4W	-11888.4	0.000 ***
R4W	M4W	-25391.2	0.000 ***

Table 3.7: Pairwise Comparisons of policies (Uplift = 7%)
Note: p-values are from paired t-tests. Significance levels are denoted by *, **, and *** for $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

for conversion.



(a) **(b)**

Figure 3.10: (a) Mean Profit by Strategy (b) Distribution prospects targeted

3.8 Managerial implications and Conclusions

Although predictive lead scoring has become a standard tool in sales and marketing operations, most approaches frame conversion as a one-time, end-of-trial outcome and rely on time-invariant usage variables, thereby limiting their capacity to support the timely qualification and prioritization of prospects. To address this limitation, we examined whether time-varying usage data, combined with weekly prediction outputs, can generate earlier and more accurate conversion signals that support targeted intervention strategies.

Drawing on 4,566 free trials from a global SaaS provider, we compared traditional machine learning models with deep learning architectures to assess how daily-granularity of usage data affect conversion prediction. The results demonstrate that temporally detailed engagement trajectories, particularly sustained activity and breadth of feature exploration, are considerably stronger predictors of conversion than prospect characteristics or time-invariant aggregates. time-varying models, especially LSTM networks, outperform traditional benchmarks by detecting reliable conversion signals early in the trial and refining them as additional usage data accumulate. To connect predictive performance with behavioral understanding, we complemented modeling with trajectory-based segmentation, which revealed distinct patterns of engagement that aligned

with both model-predicted probabilities and realized outcomes. Finally, simulation of intervention policies demonstrated that model-guided weekly outreach achieves the highest financial returns by reallocating scarce sales resources away from low-response users and toward those with the greatest incremental responsiveness.

Taken together, these findings have several implications for the management of SaaS conversion. Transitioning from time-invariant aggregates to time-varying usage data requires operational infrastructure that supports continuous data ingestion, automated scoring of temporal patterns, and workflow systems capable of triggering interventions in real time. Moreover, the superior performance of weekly, model-guided outreach highlights the importance of aligning predictive insights with operational rhythms: structured intervention cycles provide greater efficiency than continuous monitoring by synchronizing predictive accuracy with workforce capacity planning and the preparation of targeted communications. Additionally, behavioral segmentation reveals that engagement trajectories differ in both intensity and sustainability, implying that intervention design must vary accordingly. Collectively, these insights suggest that integrating temporally granular usage data into lead-scoring systems enables firms to redesign customer success operations around behavioral pattern identification, shifting trial management from a reactive approach toward a proactive system for optimizing conversion outcomes and resource allocation.

While the study demonstrates strong performance improvements, some boundary conditions must be acknowledged when interpreting these empirical findings. The analysis relies exclusively on data from a single B2B SaaS product, which may constrain the generalization of results across different software categories, market segments, or business models. to other industries or trial designs. Conversion was measured within a 90-day attribution window, excluding potential long-term effects such as subscription renewals or upselling. Moreover, while segmentation yields interpretable profiles of engagement behavior, the causal mechanisms underlying early disengagement remain to be validated through experimental or qualitative approaches.

Future research can extend these contributions in several directions. Comparative studies across industries and product types would establish whether time-varying prediction methods generalize beyond the SaaS context. Embedding real-world experiments into trial management, such as randomized targeting of marketing interventions or A/B tests of outreach timing, would provide stronger evidence on how managerial actions shape engagement trajectories, particularly for users at risk of disengagement. Beyond these extensions, advances in explainable deep learning could complement our segmentation-based approach by offering additional layers of transparency, to the interpretation of temporal prediction models and supporting their broader adoption in managerial decision-making.

In conclusion, this study advances understanding of trial-based customer acquisition by establishing time-varying as a fundamental determinant of conversion outcomes that transcends traditional prospect segmentation approaches. The integration of temporal prediction models with behavioral profiling bridges the persistent gap between predictive accuracy and managerial actionability, revealing that conversion probability emerges from evolving engagement trajectories rather than cross-sectional customer attributes. These findings contribute to the broader literature on dynamic customer behavior modeling while providing an empirically validated framework for

optimizing trial management through data-informed resource allocation strategies.

3.9 Appendix

Appendix 1

author	category	event	sessionid	source	time	user	file size	number of pages	os_platform
	document	edited	00403bdf2cf811ecee6ad89ef32c89a8		10/14/21 2:07 PM	gTLmkyO_atbCHjwi			
	tool	select object	00403bdf2cf811ecee6ad89ef32c89a8	toolbutton	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	licensing	licensing_info_active	00403bdf2cf811ecee6ad89ef32c89a8	toolbutton	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	licensing	licensing_info_expired	00403bdf2cf811ecee6ad89ef32c89a8	toolbutton	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	system		00403bdf2cf811ecee6ad89ef32c89a8	toolbutton	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			x64
	inspector	fill and stroke tab	00403bdf2cf811ecee6ad89ef32c89a8	inspector	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	inspector	fill	00403bdf2cf811ecee6ad89ef32c89a8	inspector	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	sharpen	amount	00403bdf2cf811ecee6ad89ef32c89a8	inspector	10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	tools	copy	00403bdf2cf811ecee6ad89ef32c89a8		10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	tools	paste	00403bdf2cf811ecee6ad89ef32c89a8		10/14/21 2:08 PM	gTLmkyO_atbCHjwi			
	overprint	fill overprint off	00403bdf2cf811ecee6ad89ef32c89a8	inspector	10/14/21 2:10 PM	gTLmkyO_atbCHjwi			
	document		004064512cea11ecdfbb9866da009673		10/14/21 12:26 PM	JRAE6kL1yR2kV1Bi	16717389	1	
	document		004415f62cfd11ecf484c84d7efbfdd4		10/14/21 2:42 PM	th1QSfCgb5jIE2Yh	1496691	4	
	document	edited	004415f62cfd11ecf484c84d7efbfdd4	menubar	10/14/21 2:42 PM	th1QSfCgb5jIE2Yh			
	preflight report		004415f62cfd11ecf484c84d7efbfdd4	menubar	10/14/21 2:42 PM	th1QSfCgb5jIE2Yh			
Bill H	document		00452e782cef11ecd4c2bccd36689087		10/14/21 1:02 PM	01WOdSoqISL4Xunh	814962	1	
	document		00459a9f2ce811eceeb4c8d9d2248c7b		10/14/21 12:12 PM	elnML4RLkL_BxJMf	2.98E+08	4	
medien	document		0045c3d02ce811eca306a483e74e0d43		10/14/21 12:12 PM	JDwNZhhHArMKmWGh	1.43E+08	64	
	tool	select object	0045c3d02ce811eca306a483e74e0d43	toolbutton	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			
	tool	select object	0045c3d02ce811eca306a483e74e0d43	toolbutton	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			
	inspector	position tab	0045c3d02ce811eca306a483e74e0d43	inspector	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			
	edit page boxes	define trim box	0045c3d02ce811eca306a483e74e0d43	inspector	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			
	inspector	edit page boxes	0045c3d02ce811eca306a483e74e0d43	inspector	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			
	sharpen	amount	0045c3d02ce811eca306a483e74e0d43	inspector	10/14/21 12:12 PM	JDwNZhhHArMKmWGh			

Table 3.8: Overview of data source

Appendix 2

Variable	Values
Region	Europe, North America, Asia, South America, Oceania, Africa, Unknown
Operating System	ARM-based Mac, Intel-based Mac, Intel-based, Unknown (captured from usage logs, fixed per trial)
Usage Intention	Checking (Preflight), Editing (Inspector), Fixing (Action List)

Table 3.9: Overview of prospect variables

Variable	Detail	
	Time-Varying (Daily)	**Time-Invariant (Trial)**
Activity	Binary	Ratio of active days
Dormancy	Consecutive inactive days	Final inactivity streak
Variety of actions	Distinct actions per day	Max across trial
User Count	Users per day	Max across trial
Feature use – Document	Count	Total count
Feature use – Preflight	Count	Total count
Feature use – Edit	Count	Total count
Feature use – Inspector	Count	Total count
Feature use – Tools	Count	Total count
Feature use – Others	Count	Total count

Table 3.10: Overview of usage variables

Appendix 3

Algorithm	Parameter	Specifications	Reference
Logistic regression	Penatly	L1, L2	Gattermann-Itschert and Thonemann (2022)
	Regularization	$[10^{-5}, 10^{-4}, ..., 10^{2}]$	
Random Forest	Max no. features	$\sqrt{F}, min(F, 2\sqrt{F}), min(F, 3\sqrt{F}), min(F, 4\sqrt{F})$	Gattermann-Itschert and Thonemann (2022)
	Max tree depth	[3, . . . ,. Min(F,15)]	
	Min samples per leaf	2	
XGBoost	Boosting rounds	[2, 5, 10, 20, 50, 100, 200, 500]	Janssens et al. (2022)
	Learning rate	[0.001, 0.01, 0.1, 0.2, 0.5]	
	Gamma	[0.5, 1, 1.5, 2]	

Table 3.11: Parameter setting Machine Learning algorithms

Algorithm	Parameter	Specification	Reference
1D-CNN	Filter Width	3, 4, 5	De Caigny et al. (2020); Borchert et al. (2022)
	Filters	10	
	Dropout	0.2, 0.5	
	Batch size	32	
	Hidden dimensions	100	
	L2 constraint	3	
	Lr	[0.001 - 0.01]	
	Max. epochs	15	
	Optimizer	Adamax	
LSTM	Layers	1, 2, 3, 4, 5	Sarkar and De Bruyn (2021); Beyer Díaz et al. (2024)
	Neurons	5, 20, 50, 100, 200	
	Forget bias	True, False	
	Dropout	0.2–0.35-0.5	
	Kernel initializer	RN, RU, GN, GU	
	Kernel regularizer	None, L2	
	Batch size	64	
	Early stop		
	LR	Adaptive	
	Bidirectional layers	True, False	

Table 3.12: Parameter setting Deep Learning algorithms

Appendix 4

		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Mean AUC	Std. AUC
Prospect	**Logistic Regression**	61.77%	62.79%	58.03%	63.48%	65.82%	61.70%	63.99%	65.70%	65.44%	60.19%	62.89%	0.03
	RandomForest	60.85%	64.97%	58.39%	62.65%	65.78%	60.59%	65.20%	66.94%	66.07%	61.59%	63.30%	0.03
	XGBoost	60.88%	64.19%	58.07%	63.21%	66.36%	61.35%	63.52%	67.10%	65.00%	61.35%	63.10%	0.03
Trial Level	**Logistic Regression**	77.25%	69.20%	76.23%	75.67%	75.17%	76.48%	75.82%	73.27%	72.61%	78.22%	74.99%	0.03
	RandomForest	78.02%	71.35%	76.87%	76.14%	77.09%	76.80%	75.89%	74.48%	71.65%	79.31%	75.76%	0.03
	XGBoost	77.10%	71.17%	76.53%	76.84%	76.77%	77.02%	75.38%	74.44%	70.93%	79.32%	75.55%	0.03
Daily level usage	**Logistic Regression**	77.86%	71.26%	76.07%	76.07%	76.48%	76.86%	75.32%	72.32%	72.15%	78.44%	75.28%	0.03
	RandomForest	79.00%	71.46%	75.82%	75.54%	77.23%	76.62%	75.37%	74.13%	73.10%	79.36%	75.76%	0.02
	XGBoost	79.24%	71.38%	75.92%	76.27%	76.54%	75.93%	74.54%	73.61%	72.84%	79.59%	75.59%	0.03
	CNN	75.24%	69.88%	76.29%	74.44%	75.59%	77.20%	74.71%	70.42%	71.30%	77.91%	74.30%	0.03
	LSTM	79.55%	71.65%	77.95%	76.39%	79.19%	77.28%	76.59%	75.31%	73.79%	79.64%	76.73%	0.03
Prospect + Trial level usage	**Logistic Regression**	80.07%	73.11%	76.85%	78.92%	79.48%	76.97%	79.51%	75.75%	75.41%	79.81%	77.59%	0.02
	RandomForest	80.38%	73.00%	77.88%	78.38%	78.83%	77.57%	77.44%	76.42%	74.46%	80.77%	77.51%	0.02
	XGBoost	80.00%	74.28%	77.37%	79.26%	80.58%	77.30%	79.24%	77.12%	75.20%	81.77%	78.21%	0.02
Prospect + Daily level usage	**Logistic Regression**	80.07%	74.66%	77.14%	76.88%	80.22%	76.24%	77.71%	75.28%	75.39%	79.38%	77.30%	0.02
	RandomForest	80.38%	72.86%	76.50%	76.32%	78.89%	77.29%	76.09%	75.32%	73.91%	80.13%	76.77%	0.02
	XGBoost	81.45%	74.81%	76.49%	77.50%	80.55%	77.38%	77.29%	76.59%	75.65%	81.01%	77.87%	0.02
	CNN	75.97%	70.69%	75.00%	75.10%	75.25%	76.72%	75.94%	71.14%	73.04%	78.36%	74.72%	0.02
	LSTM	81.79%	74.08%	78.33%	78.98%	79.74%	78.63%	79.98%	76.80%	74.90%	79.80%	78.30%	0.02

Table 3.13: Fine-grained 5x2 CV results

Appendix 5

Model 1	Model 2	Mean AUC Diff	Wilcoxon Stat	raw p-value	holm p-value
Prospect LR	Trial LR	0.12	0.00	0.00	0.01 **
Prospect LR	Daily LR	0.12	0.00	0.00	0.01 **
Prospect RF	Trial RF	0.12	0.00	0.00	0.01 **
Prospect RF	Daily RF	0.12	0.00	0.00	0.01 **
Prospect XGB	Trial XGB	0.12	0.00	0.00	0.01 **
Prospect XGB	Daily XGB	0.12	0.00	0.00	0.01 **

Table 3.14: Pairwise Comparisons of Prospect data vs. Usage data

Note: p-values are from paired t-tests. Significance levels are denoted by *, **, and *** for $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Appendix 6

Model 1	Model 2	Mean AUC Diff	Wilcoxon Stat	raw p-value	Holm p-value
Trial LR	Daily LR	0.00	18.00	0.38	1.00
Trial RF	Daily RF	0.00	24.00	0.77	1.00
Trial XGB	Daily XGB	0.00	23.00	0.70	1.00
Daily LR	Daily LSTM	0.01	0.00	0.00	0.02 **
Daily RF	Daily LSTM	0.01	0.00	0.00	0.02 **
Daily XGB	Daily LSTM	0.01	0.00	0.00	0.02 **
Daily LR	Daily CNN	-0.01	3.00	0.01	0.06
Daily RF	Daily CNN	-0.01	3.00	0.01	0.06
Daily XGB	Daily CNN	-0.01	7.00	0.04	0.15

Table 3.15: Pairwise Comparisons of trial aggregated vs. daily usage data

Note: p-values are from paired t-tests. Significance levels are denoted by *, **, and *** for $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Appendix 7

Figure 3.11 presents a sensitivity analysis of profitability across uplift assumptions $u \in \{0.05, 0.07, 0.09\}$, further confirming the robustness of the simulation findings. Across all uplift levels, the Model-Based (4 Weeks) strategy consistently yields the highest average profit within Quartile 4, illustrating its effectiveness in targeting high-propensity users and timing interventions optimally. The combination of predictive targeting and temporal flexibility remains particularly effective for conversion-likely prospect. In contrast, users in Quartile 1 exhibit smaller marginal gains. These findings reinforce the value of adaptive, model-based strategies in maintaining performance across varying treatment effect assumptions.

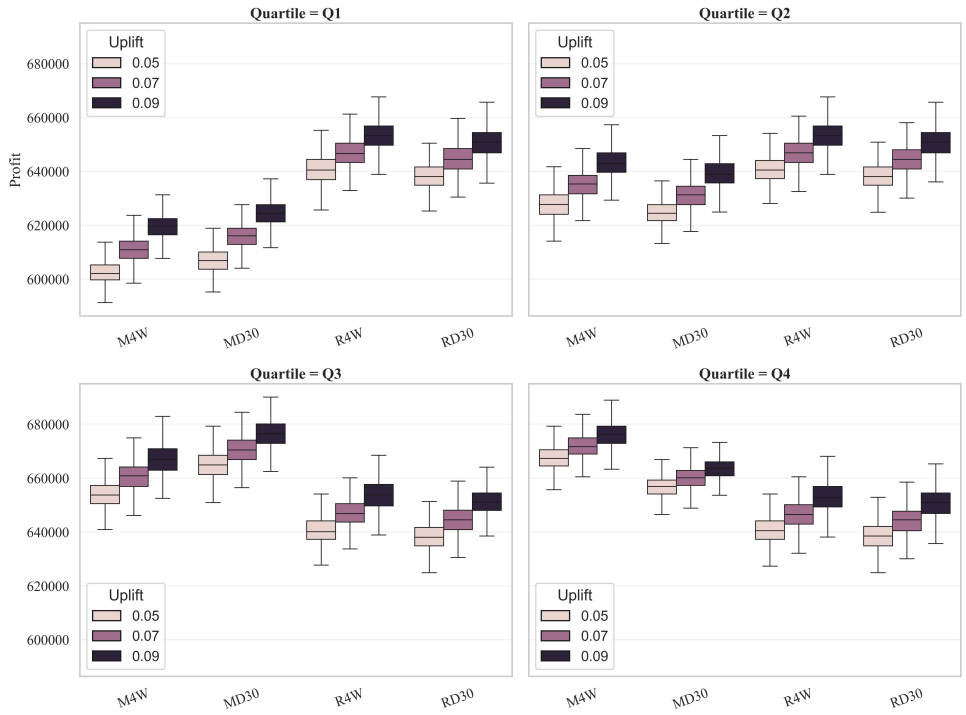


Figure 3.11: Profit Comparison: Random vs Model Targeting by Predicted Probability Quartile

Appendix 8

Figure 3.12 indicates that, although all policies contact a comparable number of prospects under the fixed budget constraint, the composition of outcomes varies considerably across strategies. In the high-propensity quartiles (Q3 and Q4), model-based approaches yield substantially more incremental conversions, with the dynamic weekly strategy (M4W) producing the highest number of new subscriptions relative to MD30 and the random benchmarks. By contrast, random policies allocate a greater proportion of contacts to prospects who either convert without intervention or fail to respond, thereby reducing their effectiveness.

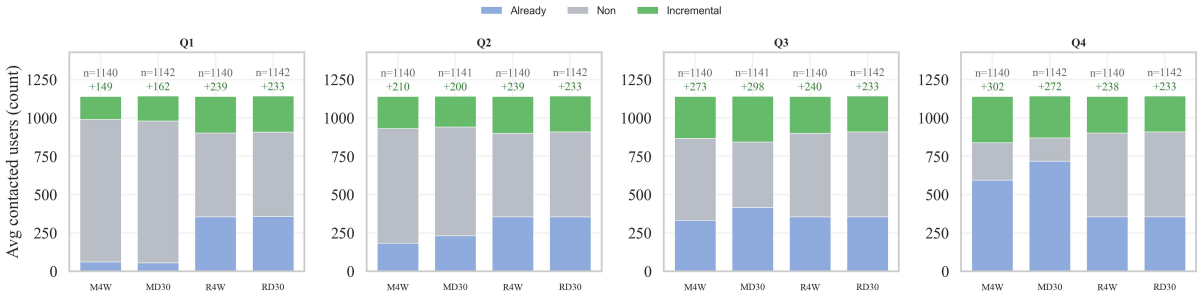


Figure 3.12: Outcome Composition Among Contacted prospects per policy (Uplift = 7%)

Appendix 9

Uplift	Scenario 1	Scenario 2	Mean Diff	p-value
0.05	RD30	MD30	-18150	0.000 ***
0.05	RD30	R4W	-2090.4	0.000 ***
0.05	RD30	M4W	-28956.8	0.000 ***
0.05	MD30	R4W	16059.6	0.000 ***
0.05	MD30	M4W	-10806.8	0.000 ***
0.05	R4W	M4W	-26866.4	0.000 ***
0.09	RD30	MD30	-12764.4	0.000 ***
0.09	RD30	R4W	-2276.4	0.000 ***
0.09	RD30	M4W	-25480	0.000 ***
0.09	MD30	R4W	10488	0.000 ***
0.09	MD30	M4W	-12715.6	0.000 ***
0.09	R4W	M4W	-23203.6	0.000 ***

Table 3.16: Pairwise Comparisons of policies (Uplift = 5% and 9%)

Note: p-values are from paired t-tests. Significance levels are denoted by *, **, and *** for $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

References

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1):80–98.

Ascarza, E. and Hardie, B. G. (2013). A joint model of usage and churn in contractual settings. *Marketing Science*, 32(4):570–590.

Baumann, A., Haupt, J., Gebert, F., and Lessmann, S. (2018). Changing perspectives: Using graph metrics to predict purchase probabilities. *Expert Systems with Applications*, 94:137–148.

Beyer Díaz, S., Coussement, K., De Caigny, A., Pérez, L. F., and Creemers, S. (2024). Do the US president's tweets better predict oil prices? An empirical examination using long short-term memory networks. *International Journal of Production Research*, 62(6):2158–2175.

Bolton, R. N. and Lemon, K. N. (1999). A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction. *Source: Journal of Marketing Research*, 36(2):171–186.

Borchert, P., Coussement, K., De Caigny, A., and De Weerdt, J. (2022). Extending business failure prediction models with textual website content using deep learning. *European Journal of Operational Research*.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32.

Chaudhuri, N., Gupta, G., Vamsi, V., and Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*, 149.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Cheng, H. K., Li, S., and Liu, Y. (2015). Optimal software free trial strategy: Limited version, time-locked, or hybrid? *Production and Operations Management*, 24(3):504–517.

Chou, P., Chuang, H. H. C., Chou, Y. C., and Liang, T. P. (2022). Predictive analytics for customer repurchase: Interdisciplinary integration of buy till you die modeling and machine learning. *European Journal of Operational Research*, 296(2):635–651.

Danaher, P. J., Danaher, T. S., Smith, M. S., and Loaiza-Maya, R. (2020). Advertising Effectiveness for Multiple Retailer-Brands in a Multimedia and Multichannel Environment. *Journal of Marketing Research*, 57(3):445–467.

Datta, H., Foubert, B., and Van Heerde, H. J. (2015). The challenge of retaining customers acquired with free trials. *Journal of Marketing Research*, 52(2):217–234.

De Caigny, A., Coussement, K., De Bock, K. W., and Lessmann, S. (2020). Incorporating textual information in customer churn prediction models based on a convolutional neural network. *International Journal of Forecasting*, 36(4):1563–1578.

D'Haen, J., Van Den Poel, D., and Thorleuchter, D. (2013). Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique. *Expert Systems with Applications*, 40(6):2007–2012.

Gattermann-Itschert, T. and Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*, 107:134–147.

Goic, M., Rojas, A., and Saavedra, I. (2021). The Effectiveness of Triggered Email Marketing in Addressing Browse Abandonments. *Journal of Interactive Marketing*, 55:118–145.

González-Flores, L., Rubiano-Moreno, J., and Sosa-Gómez, G. (2025). The relevance of lead prioritization: a B2B lead scoring model based on machine learning. *Frontiers in Artificial Intelligence*, 8.

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, volume 2. New York: springer.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural computation*, 8(9):1735–1780.

Janssens, B., Bogaert, M., Bagué, A., and Van den Poel, D. (2022). B2Boost: instance-dependent profit-driven modelling of B2B churn. *Annals of Operations Research*.

Järvinen, J. and Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management*, 54:164–175.

Khan, Z. Y. and Niu, Z. (2021). CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications*, 170.

Kiranyaz, S., Ince, T., Hamila, R., and Gabbouj, M. (2015). Convolutional Neural Networks for Patient-Specific ECG Classification. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

Lemon, K. N., White, T. B., and Winer, R. S. (2002). Dynamic Customer Relationship Management: Incorporating Future Considerations into the Service Retention Decision. *Journal of Marketing*, 66(1):1–14.

Li, B. and Kumar, S. (2022). Managing Software-as-a-Service: Pricing and operations. *Production and Operations Management*, 31(6):2588–2608.

Li, H. (2022). Converting free users to paid subscribers in the SaaS context: The impact of marketing touchpoints, message content, and usage. *Production and Operations Management*, 31(5):2185–2203.

Lin, J., Choi, T. M., and Kuo, Y. H. (2024). Online Retailing Operations: Is It Wise to Offer Free Sample Trial and Money-Back Guarantee Together? *Production and Operations Management*, 33(11):2158–2176.

Loukis, E., Janssen, M., and Mintchev, I. (2019). Determinants of software-as-a-service benefits and impact on firm performance. *Decision Support Systems*, 117:38–47.

Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., and Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, 281(3):588–596.

Mehra, A. and Saha, R. L. (2018). Utilizing Public Betas and Free Trials to Launch a Software Product. *Production and Operations Management*, 27(11):2025–2037.

Meire, M., Ballings, M., and Van den Poel, D. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. *Decision Support Systems*, 104:26–37.

Moe, W. W. and Fader, P. S. (2004). Dynamic Conversion Behavior at E-Commerce Sites. *Management Science*, 50(3):326–335.

Naik, P. A. and Raman, K. (2003). Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research*, 40(4):375–388.

Oliveira, T., Martins, R., Sarker, S., Thomas, M., and Popovič, A. (2019). Understanding SaaS adoption: The moderating impact of the environment context. *International Journal of Information Management*, 49:1–12.

Rafieian, O. and Yoganarasimhan, H. (2021). Targeting and privacy in mobile advertising. *Marketing Science*, 40(2):193–218.

Reiners, R., Haubitz, C. B., and Thonemann, U. W. (2025). Lead Time Prediction for Inventory Optimization With Machine Learning. *Production and Operations Management*.

Sahni, N. S., Wheeler, S. C., and Chintagunta, P. (2018). Personalization in Email Marketing. *Marketing Science*, 37(2):236–258.

Sakar, C. O., Polat, S. O., Katircioglu, M., and Kastro, Y. (2019). Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications*, 31(10):6893–6908.

Sanchez Ramirez, J., Coussement, K., De Caigny, A., Benoit, F., Waardenburg, L., and Guliyev, E. (2024). Incorporating Usage Data for B2B Churn Prediction Modeling. *Industrial Marketing Management*, 120:191–205.

Sarkar, M. and De Bruyn, A. (2021). LSTM Response Models for Direct Marketing Analytics: Replacing Feature Engineering with Deep Learning. *Journal of Interactive Marketing*, 53:80–95.

Seethamraju, R. (2015). Adoption of Software as a Service (SaaS) Enterprise Resource Planning (ERP) Systems in Small and Medium Sized Enterprises (SMEs). *Information Systems Frontiers*, 17(3):475–492.

Somanchi, S., Kanuri, V. K., and Telang, R. (2025). Mitigating Churn After Online Financial Fraud: The Value of Blame Attribution. *Production and Operations Management*, pages 1–19.

Sriram, S., Chintagunta, P. K., and Manchanda, P. (2015). Service quality variability and termination behavior. *Management Science*, 61(11):2739–2759.

Wang, T., Jin, F., Hu, Y. J., Feng, L., and Cheng, Y. (2024). Making Early and Accurate Deep Learning Predictions to Help Disadvantaged Individuals in Medical Crowdfunding. *Production and Operations Management*.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics bulletin*, 1(6):80–83.

Wu, M., Andreev, P., Benyoucef, M., and Hood, D. (2024). Unlocking B2B buyer intentions to purchase: Conceptualizing and validating inside sales purchases. *Decision Support Systems*, 179.

Yoganarasimhan, H., Barzegary, E., and Pani, A. (2023). Design and Evaluation of Optimal Free Trials. *Management Science*, 69(6):3220–3240.

Zantedeschi, D., Feit, E. M. D., and Bradlow, E. T. (2017). Measuring multichannel advertising response. *Management Science*, 63(8):2706–2728.

Chapter 4

Mirroring in AI Design: Challenges and Organizational Impacts

Chapter 4
Mirroring in AI Design: Challenges and Organizational Impacts

Abstract

As organizations increasingly integrate AI-driven systems, understanding their impact on business structures becomes more relevant. AI technologies frequently introduce complexity that challenges traditional workflows, decisions, and roles. This paper examines the alignment between technological (AI) and organizational designs to understand how AI integrates with current business structures and how those structures adapt over time. This paper presents research on the relationship between these designs, drawing on a longitudinal study spanning four years of an AI system for customer retention at a European B2B software provider. Through the analysis of various data sources, including interviews with end-users and stakeholders, direct observation, and email correspondence, we examined both the practical use and reflective insights provided by designers and organizational experts regarding the development of the system. Framed by the "mirroring hypothesis", our study reveals that a narrow focus on organizational and product design fails to capture the complex dynamics of AI development. We explore how mirroring unfolds across phases, influenced by diverse stakeholders whose involvement configures the balance between modularity and integration in technological and organizational designs. Additionally, we explore how the technological design can unintentionally reflect disconnections within the organizational structures, leading to what we call "distorting the mirror". Moreover, we argue that AI development, as a co-creation process, is not neutral; it actively determines what is mirrored, emphasizing the complex, cumulative nature of mirroring in AI contexts. This study offers guidelines for organizations to facilitate co-creation processes during AI development in multi-stakeholder settings.

Keywords: Artificial intelligence, integration, modularity, mirroring hypothesis, co-creation

4.1 Introduction

The ability of artificial intelligence (AI) to continuously learn from its environment through machine learning algorithms has encouraged recent calls for research beyond development and use (e.g. Bailey and Barley (2020)). These calls emphasize the need for a (re)turn to a relational perspective on technology in both organizational and information systems (IS) research, which allows us to understand technology "not as stable entities, but as a set of evolving relations" (Bailey et al., 2022). This means that AI is a process of continuous co-creation (Van den Broek et al., 2021) that requires the involvement of various organizational actors. The development and implementation of AI in organizations is, thus, likely to alter and reshape technological and organizational boundaries that have traditionally been put in place (Bailey and Barley, 2020). However, how technological and organizational designs are related to each other in the case of AI development remains under-explored.

A useful lens to study this relationship is what is commonly understood in IS literature as the 'mirroring hypothesis.' The mirroring hypothesis posits that an organizational structure, including its communication and workflows, mirrors and is mirrored in the architecture of the product it develops (e.g., Henderson and Clark (1990); Cabigiosu and Camuffo (2012); MacCormack et al. (2012); Colfer and Baldwin (2016); Sorkun and Furlan (2017)). In other words, as described by Sanchez and Mahoney (1996) products with tightly integrated components tend to be developed by organizations with tightly integrated teams, and modular products are more likely to be developed by organizations with decentralized or modular structures. An example of the former are custom-built applications, such as enterprise resource planning (ERP) systems, which optimize performance by incorporating highly integrated components. This approach demands severe managerial coordination, as changes to one part of the system often require adjustments to others. In contrast, modular software architectures, such as SaaS solutions, allow companies to mix and match different modules, allowing vendors to develop components independently within a loosely coordinated structure.

The mirroring hypothesis posits that a structural alignment exists between the architecture of a product and the organization creating it (Henderson and Clark, 1990). Subsequent research, expanded the hypothesis to platforms, ecosystems, and open-source communities, examining varying degrees of mirroring within firms, across firms, and in open collaboration. However, recent studies highlight exceptions where organizations "break the mirror," leading to coordination challenges and inefficiencies. While the mirroring hypothesis is a widely accepted theoretical lens, the study of AI and its evolving complexity offers new opportunities to further unpack this hypothesis. This study extends the mirroring hypothesis to AI system development, which consists of modular components that must be tightly integrated for the system to function as intended. Yet how organizations cope with the tension between modularity and integration in AI development, and the resulting implications for organizational structure, remain unasked.

For this reason, we have conducted a longitudinal, qualitative study from November 2019 to November 2023, on the development and implementation of an AI system for customer retention at a European B2B software provider (pseudonym "SoftCo"). This case study is characterized

by a rich collection of data sources, including 27 interviews with end-users and stakeholders, direct observation over a two-day period, email correspondence spanning two years, and meeting minutes from a three-year period. By looking into three key phases in AI development – data understanding and preparation, model definition, and tool integration – we unpack how different technological and organizational requirements trigger a sequence of mirroring activities. Moreover, we highlight how, in this mirroring process, the tool ended up mirroring the disconnect between the SoftCo and the end user. This resulted in what we call “distorting the mirror", in which the system mirrors a misaligned or distorted version of the organizational design it was intended to represent.

Approaching mirroring as a dynamic process, we provide new insights into the mirroring hypothesis literature (Henderson and Clark, 1990; Cabigiosu and Camuffo, 2012; MacCormack et al., 2012; Colfer and Baldwin, 2016; Sorkun and Furlan, 2017), exploring details about how it unfolds and operates. This study shows that understanding what is being mirrored in the case of AI development extends beyond organizational and product design to also incorporate end users. Moreover, the study emphasizes that mirroring is accomplished through multiple, cumulative phases involving different stakeholders and requirements. Our study also offers contributions to the emerging literature on AI development as a co-creation process by emphasizing its subjective nature and the critical role of stakeholder interactions within the co-creation framework. This underscores the importance of carefully selecting the participants in the co-creation process to ensure that the final AI tool effectively meets its intended objectives and mirrors the requirements of all relevant stakeholders.

4.2 Theoretical Framing

4.2.1 Modularity and AI

Modularity is commonly understood in IS and organizational design literature as an attribute of complex systems that enables the division of work into simpler tasks or modules by minimizing dependencies between them and maximizing interdependence within each module (Campagnolo and Camuffo, 2010). As such, the concept is often used to better understand complex system design (Baldwin and Clark, 2000). For example, building on the principles of complex systems evolution presented by Holland (1992), Baldwin and Clark (2000) developed a theory of design evolution using “modular operators" connected as a network of interdependent modules. They explore how different configurations of these interactions along a network can be combined to improve the system’s performance (Langlois, 1992).

Generally, two types of complex design processes have been central to the study of modularity. First, organizational design, which can take multiple configurations tailored to the products and needs of companies, ranging from vertical and horizontal designs to modular or virtual organizations. Second, product or technology design, conceptualized as a creative activity that requires the systematic coordination and standardization of various elements to achieve predefined objectives. Modularity in product design refers to the degree to which a system’s components

can be separated and recombined. A modular product, therefore, is designed in such a way that individual components (or modules) are developed independently but still work together as a whole (Ravasi and Stigliani, 2012).

Scholars emphasize that organizational and product design often co-occur (e.g., Baldwin (2023); Puranam et al. (2012)). For instance,Colfer and Baldwin (2016) in their empirical study on the aircraft and automotive industries, define technical architecture as the allocation of technical functions to system components and their interdependencies, capturing the principle of "What depends on what." Additionally, they analyse the division of labour in the development of this architecture, addressing the question of "Who does what." A correspondence scheme between these two dimensions can be established, wherein the modules of the technical architecture are assigned to individuals or teams. This can occur in two ways: either product design dictates team structures, shaping the organizational division of labour, or conversely, organizational structures influence the configuration of the technical architecture.

Building on these insights, the concept of modularity can be extended beyond traditional product and organizational design to the development of AI. To introduce the term AI, a first referent is the 'Imitation game' proposed by Turing (1950). He introduced the concept of machine intelligence by assessing a machine's capacity to demonstrate intelligent behaviour indistinguishable from that of a human. Later, additional studies provided formal definitions of AI, suggesting that machines could be programmed to solve well-defined problems, while also having the potential for self-improvement (McCarthy et al., 1955; Nilsson, 1980). This involved interpreting external data, learning from it, and applying the insights gained to address specific problems (Haenlein and Kaplan, 2019).

Although AI has been present for decades, its benefits and practical applications for organizations have only recently become evident with advancements in data collection and computational power. Those advancements enabled the development and implementation of more sophisticated algorithms capable of addressing complex problems, including the prediction of customer engagement or sales (Ma et al., 2024; Chen et al., 2024), as well as evaluating the impact of marketing campaigns on customer traffic (Wang et al., 2022). With advancements in so-called "Machine Learning" (ML), modularity has become an integral aspect of these developments, particularly within the algorithms themselves, enabling multiple models to work as modular components that can be stacked or combined to enhance performance and adaptability. For example, models such as Neural Networks, introduced by Nobel Prize in physicsHopfield (1982), act as modular components within a larger system. Each neuron is designed to learn a specific aspect of the data representation, allowing these modules to be stacked progressively to tackle more complex learning tasks.

Beyond its role in the technical design of AI models, modularity is also present in the organizational processes of AI development, typically structured into well-defined stages. For example, Shrestha et al. (2021) describe the stages of business understanding, data understanding, data preparation, modeling, evaluation, and implementation, which operate as independent modules with specific purposes, while collectively contributing to the overall model development. This modular framework allows businesses to scale their AI systems efficiently, reconfigure

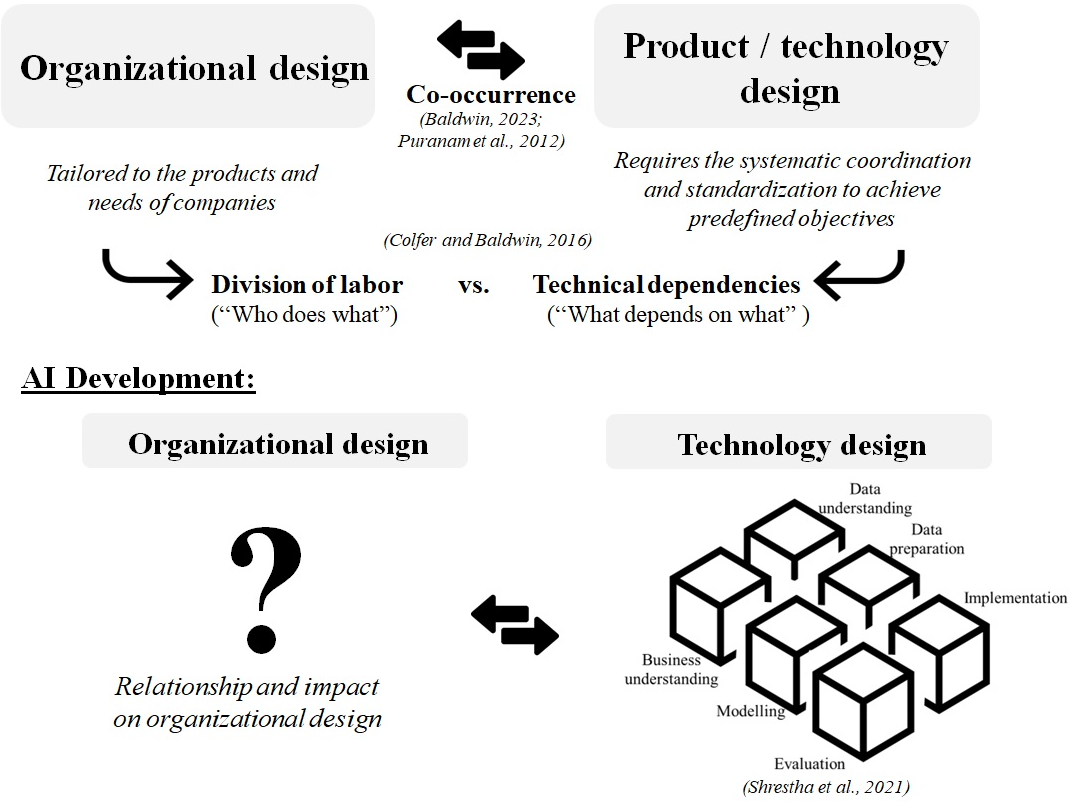


Figure 4.1: Technical architecture and the division of labour in AI development

individual components without requiring a complete overhaul, and adapt to evolving technological and operational requirements (Sjödin et al., 2021).

Moreover, one of the key insights of previous research on modularity that currently remains under-explored in relation to AI development, is that organizational and product design are typically interrelated. For example, Colfer and Baldwin (2010) examine the coordination between technical architecture and the division of labour. They argue that through such coordination and relationships, one component can increasingly reflect and correspond to the structure of the other. This has been defined in literature as the mirroring hypothesis (Henderson and Clark, 1990; Colfer and Baldwin, 2016). Figure 4.1 summarizes the interconnection between technical architecture and the division of labour particularly for an AI development processes.

4.2.2 Mirroring Hypothesis

The first to coin what we nowadays call the "mirroring hypothesis" were Henderson and Clark (1990) in their work on product innovation. They argue that there is a structural alignment between the design of a product and the organization that creates it. In other words, an organizational structure, including its communication and workflows, mirrors and is mirrored in the architecture of the product it develops. As such, products with tightly integrated components tend to be developed by organizations with tightly integrated teams and communication networks. On the other hand, modular products that consist of independent components that can be developed separately are more likely to be developed by organizations with decentralized or

modular structures. Accordingly, Henderson and Clark (1990) suggest that modular product designs enable more flexible and decentralized organizational structures.

Particularly in IS, the mirroring hypothesis has received much attention, especially between 2010 and 2017 (Cabigiosu and Camuffo, 2012; MacCormack et al., 2012; Colfer and Baldwin, 2016). With the emergence of new organizational forms in subsequent years, such as platforms, ecosystems, and open-source communities, where modularity was understood to play a crucial role (Baldwin, 2023; Colfer and Baldwin, 2016)), the hypothesis has been further developed to include how technology shapes organizational structures in modern contexts by identifying different degrees of mirroring and comparing them with traditional close-coordinated organizational settings (Baldwin, 2023). For example, Colfer and Baldwin (2016) cover three types of settings where mirroring can occur – i.e., within firms, across firms, and open collaborative environments – and analyse the prevalence and extent of mirroring in each setting. They first discuss the mirroring of technical dependencies within firms, for which mirroring is the most intuitive pattern. Then, they explore across-boundaries mirroring by studying buyer-supplier relationships and their potential interdependences. Finally, they consider open collaborative environments, particularly in technology-related contexts where participants work independently. According to these three types of settings, Colfer and Baldwin (2016) found that, while mirroring is a common pattern, it is not universal within firms or industries, and its superiority depends on how clearly organizational and knowledge boundaries are defined, i.e., the better defined the boundaries, the higher the degree of mirroring.

Recent research has also explored how technology allows groups to deviate from traditional mirroring structures observed in firms (Colfer and Baldwin, 2010). In these settings, organizational structures can lead to more complex organizational dynamics than the mirroring hypothesis. More specifically, it can lead to instances where companies strategically “break the mirror" (Colfer and Baldwin, 2010), in which there is a deliberate or accidental misalignment between the design of a product and the organizational structure responsible for creating it. In general, exceptions to the mirroring hypothesis, or mirror-breaking, arise in two scenarios: either a highly connected group creating a modular product, or independent contributors working on an interdependent system (Colfer and Baldwin, 2010). Such exceptions can lead to multiple challenges. For example, when the structure of a product does not align with the organization’s structure, it can create coordination difficulties (Colfer and Baldwin, 2016)if different teams are responsible for independent modules but the product requires integration, poor communication or lack of collaboration can cause inefficiencies, integration issues, and delays (Sosa et al., 2004). Moreover, in complex environments where products are continuously evolving, the rigid alignment between organizational and product structures may be inefficient or undesirable.

It is particularly considering this last challenge that the study of AI and its evolving complexity offers an interesting perspective to further unpack the mirroring hypothesis. On one hand, AI systems are built from distinct, modular components—such as data preprocessing, model training, and evaluation. On the other hand, these systems require that their various layers, such as input, processing, and output (Von Krogh, 2018)operate in tight synchronization. Since, changes in one component often affect the entire system, developers must integrate properly the existing

modules. Examining how this tension between modularity and integration in AI design mirrors organizational structures may reveal, for example, how the architecture of an AI system both influences and is influenced by the organizational framework overseeing its development. It is for this reason that we ask: *How do organizations cope with the tension between modularity and integration in AI development, and with what consequences for organizational structure?*

4.3 Methods

4.3.1 Research context

This study was conducted at a European B2B software provider (pseudonym "SoftCo") with over 30 years of experience in the market. On their portfolio of products, they provide different types of software related to PDF validation and process automation, which they sell and distribute to the final customers mainly through their channel partners, composed of distributors and resellers across the globe. Currently, they sell their products in two types of formats. First, and at present the most used, consists of a perpetual license where the customer acquires the software and has the choice to pay a yearly maintenance contract. Second, they offer a subscription-based system where customers opt for monthly or yearly subscriptions and always have access to updates, support, and training. Nevertheless, when the subscription is cancelled, the customers do not have access to the software any longer. Maintenance contracts are mandatory for the first year of the perpetual license and optional for the subsequent years. The maintenance contract provides (1) access to exclusive content on the training platform, (2) direct support from partners and the support portal of the company, and (3) access to updated versions of the software with new features and functionalities.

While this contract offers additional benefits, the percentage of customers who did not renew their contracts reached 20% in 2019, which is notably high compared to the industry average of 8% (Meyer et al., 2023), particularly since maintenance contracts constituted the largest portion of recurring revenue. This rate will henceforth be referred to as the churn rate. This triggered the company to look into the development of a customer retention tool (also referred as churn prediction tool throughout this paper) using machine learning to predict the likeliness for existing customers to churn. By developing this tool, the company expected to identify customers with a high propensity to churn and provide this information to channel partners, allowing them to prioritize retention efforts.

This project officially started in January 2020, after a series of preliminary meetings with a team of technical experts who would oversee this development. One data scientist played a pivotal role as the lead technical expert, leading data collection, engaging with business experts, and conducting experiments. Meanwhile, two additional specialists provided guidance on state-of-the-art methodologies and best practices. This project was the first AI development in the company and involved different business experts from diverse areas of the company, specifically customer success, business operations, product, and support. The lifecycle of the project included several key stages, including a year and a half dedicated to data understanding and cleaning,

followed by six months focused on experimental setup and tool integration. After two years of collaborative effort, the tool was deployed into the internal Customer Relationship Management (CRM) Salesforce portal, providing access to the designated end users: the company's so-called "channel partners". The prediction results were displayed on the CRM portal, highlighting the top 10 customers with the highest likelihood of churn for each channel partner's respective customer base, which should trigger the channel partners to adjust their retention strategies to prevent those customers from cancelling their maintenance contract.

4.3.2 Data Collection

Data types	Amount	Description
Primary data		
Physical observation	2 days	Data on the notes of the 'Channel event' organized by SoftCo where channel partners took part in various activities and discussions focused on technical aspects, sales, and customer success management. The event in April 2023.
First round of interviews	17 (10 h)	Data on the understanding of business practices of the channel partners, their approach to customer retention and point of view. The interviews were conducted from May 2023 to September 2023 with partners from diverse regions of the world.
Second round of interviews	10 (10 h)	Data on the development of the churn prediction tool, as well as SoftCo's approach and current practices regarding customer churn. The data collection took place between October and November 2023, involving business experts including the Head of Customer Success (1), Customer Success team lead (1), Customer Success Managers (3), Head of Product (1), Head of Sales (1), Product Manager (1), Software Architect (1), and Lead Technical Expert (1) at SoftCo.
Secondary data		
Meeting minutes	48 (87 pg.)	Data on the summary of the meetings between the technical expert team. Collected from November 2019 to September 2022
Email interactions		Data on interactions between business experts and technical experts related to details about the data understanding and consistency. The data includes emails between November 2019 and November 2021

Table 4.1: Summary data types

To address our research question, we adopted a qualitative research design, more specifically an in-depth explanatory case study (Yin, 2009). Our study draws on various data sources (see Table 4.1 for a complete list). First, the author team had full access to all internal communication regarding the development of the churn prediction system from its inception. Digging into email

interactions between technical and business experts, along with a series of meeting minutes among technical experts, allowed us to retrospectively reconstruct some of the initial steps and challenges encountered in the development process. We also had direct access to the lead technical expert, who helped us to understand the intricate details of the AI system and who helped us reconstruct a detailed map of the various stages involved in the development and implementation process.

Moreover, the lead technical expert helped us to identify the key stakeholders involved throughout the process, which led to a total of 27 interviews (lasting between 17 and 85 minutes) with, for example, business experts, technical experts, and channel partners. Due to the complex and multi-stakeholder nature of the AI project, we conducted the interviews in multiple rounds. The first round of interviews, which took place between May and September 2023, was kicked off by the first author's participant observations at a two-day 'channel event' involving representatives of the intended end-users of the churn prediction tool. After making the initial connections at this event, we interviewed the channel partners (in total 17 interviews). Our interviews with channel partners were aimed at understanding how the AI system was used and gaining retrospective insights from the end-users into the implementation process. During these interviews, we were also particularly interested in finding out how the end-users perceived the role of churn in their work and how they related this to the predictions generated by the AI system. We asked questions such as "Regarding customer retention, could you please describe the specific actions that you or your company take to prevent customer churn?". In the second round of interviews, between October and November 2023, we interviewed business and technical experts (in total 10 interviews). In these interviews, we intended to uncover retrospective reflections on the development and implementation process from both the designer and the organizational side. During these interviews, we asked questions such as *"In your opinion, what improvements can SoftCo make to this tool to better assist partners in customer retention tasks?"* or *"What factors do you believe this [AI] tool should take into account when predicting customer churn?"*. For a comprehensive list of the questions, the full semi-structured interview guide is included in Appendix 4.7.

All interviews were conducted in either English, Spanish, or French and were recorded and transcribed verbatim. Combining both interview rounds offered us a holistic perspective on the development and implementation of the churn prediction tool, from the developer, organizational, and end-user side.

4.3.3 Data Analysis

The data analysis was performed iteratively and concurrently with the data collection process. First, we created a chronological presentation (Miles and Huberman, 1994) of the development and implementation process of the churn prediction model. Through this chronological presentation, we realized the strong role of the lead technical expert and the business experts from the hosting organization in determining the general understanding of churn and, therefore, the underlying rationale of the AI system. In constructing the initial chronological narrative, we also noticed that the development team faced various challenges in aligning the design choices of the AI system with existing organizational processes. We grouped these challenges into three generic

categories: (1) data, (2) modelling, and (3) implementation.

We then performed several rounds of open coding by reading the transcripts of the interviews and highlighting relevant comments that could lead to identifying emerging topics. This process was conducted separately for each round of interviews to maintain the context of shared topics as, for instance, channel partners discussed topics relevant to each other but not necessarily to business experts. In interviews with partners, we identified general topics such as the dynamics of the company-partner relationship, the perceived value of the products offered, and their relationships with end customers. In the interviews with the business experts and the technical experts, we identified general topics such as interactions with partners, approaches to customer retention, and the development of the churn prediction tool.

After the initial rounds of open coding, we performed more targeted rounds of coding in which we identified more specific codes for each topic. For example, for the channel partners, we included detailed codes such as the tools they used for communication with the company and customers, issues related to information access, customer feedback on the software, and opinions on the relevance of churn management. For the interviews with the technical and business experts, we identified detailed codes that provided a clearer understanding of the decision-making process, initial project expectations, key challenges, and project outcomes.

By grouping these more targeted codes according to the three generic categories that we initially identified, we could distil three more specific phases of AI development: (1) data understanding and preparation, (2) model definition, and (3) tool integration. Moreover, through the coding process, we noticed that the understanding of churn of the end-users varied from how the technical experts and business experts established it in the development process. We presented these empirical details at a conference and were pointed by our peers at the mirroring hypothesis as a theoretical interpretation of our observed dynamics. Triggered by this suggestion, we used the literature on the mirroring hypothesis to better understand the dynamics and outcomes of the three generic phases. We first identified the specific activities performed at each of the three phases. Phase 1 included data understanding, correction, selection, and preparation. Phase 2 involved feature selection, feature revision, model evaluation, and agreeing on the final model. Phase 3 encompassed the presentation and use of the AI tool.

We represented these stages and activities in an organizational ties matrix (Colfer and Baldwin, 2016) which combines a technical dependency matrix with the association of teams involved in each activity, since the mirroring hypothesis predicts that, given a dependency between certain tasks, the actors involved in those tasks should share one or more explicit organizational ties that enable them to coordinate their actions (see Figure 4.2).

The analysis of each phase revealed its specific outcomes in relation to mirroring, partial mirroring, or mirror-breaking. This, finally, helped us explaining the eventual misalignment between the underlying rationale of the churn prediction tool and the understanding of churn as specified by the end-users. Below, we use the identified phases to explain the mirroring activities, reflecting how the organization coped with the tension between modularity and integration, involved in the development and implementation of AI in practice.

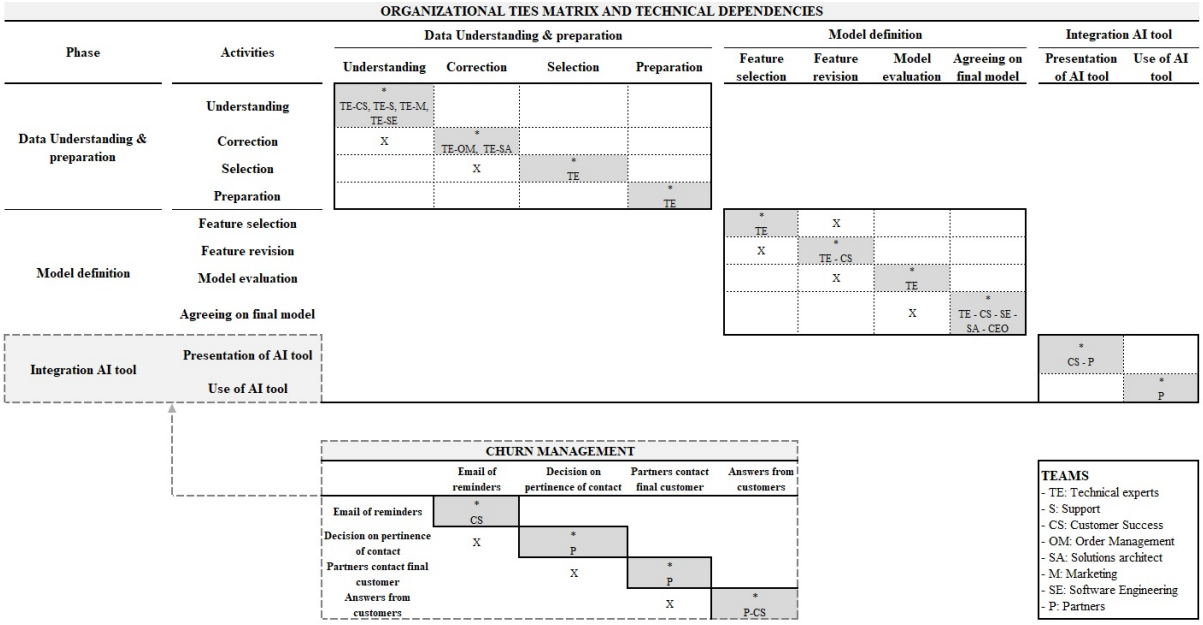


Figure 4.2: Organizational Tias and Technical dependencies

4.4 Findings

Given the business context of SoftCo, the general purpose of the project was to develop an AI tool to predict customer churn. For the customer success team, which oversees recurring revenue, the pertinence of such a tool and its added value was to give an integrated perspective on customer behaviour, which they described as a transition towards the *"recurrent revenue world"*:

> "We were learning that as we were moving more and more towards the recurring revenue world, we needed new ways of thinking, and new strategies to live in that world and churn is a very big indicator ... So, we needed to make sure that we understood this, that we knew how many customers we were losing, what the impact was of the customers we are losing and how we could avoid losing customers." – Head of Customer Success

In other words, the head of customer success who was the initiator of the project had clear expectations regarding the tool's value proposition. These expectations consisted of quantifying, identifying, and understanding churn-risk populations to later take preventive actions aligned with the insights obtained from the model:

> "I really was hoping that by identifying the high-potential churners we could work on those and give them tailor-made propositions so that they would renew and not churn. Yeah, that was really the basic thinking of this churn prediction tool and because we have 6000 contracts, we cannot focus on all of them together. Yeah, that's too much... Well, the goal was first of all give this to the channel partners, so they know which customers are more likely to churn in their region... at least we should use it internally for us to know which customers we want to focus on." – Head of Customer Success

To develop the churn prediction tool, particular modular elements of the product design, such as the time window, types of data to use, and the model design steps to follow, had to be first established by the technical experts. However, to do this, the technical experts quickly figured that they needed business input, which led to various meetings with business experts. For example, the head of customer success provided a general context explaining the history, structure, industry, products, business model, and goals of the company which helped the technical experts to better understand SoftCo's goals for developing the tool. Also, the lead technical expert had several meetings with different business experts to understand the specific operation of each area. He met the head of product, marketing, product owners, sales, and development teams. Consequently, the lead technical expert gained a clearer, more integrated grasp of what kind of tool could contribute to the interests of the business experts. This understanding was key, given the fact that this information would allow him to structure the subsequent steps of the project:

> "So, the recurring revenue model motivates companies to really find a churn model that will help them retain their customers. And when we started, they didn't know what was possible, but they wanted to test, so we set up the project... Yeah, because it's alarming your revenue stream is ... highly skewed to recurring revenue and every single month, [churn] 20% is a lot of money. If they could just bring back 5%, it could help them a lot financially. So, the financial motive was very visible." – Lead technical expert

Moreover, during the project definition process, the lead technical expert sat down with business experts from different areas within SoftCo, such as support, business operations, and customer success, to gain valuable insights that, based on their industry knowledge, could help predict customer churn. By leveraging these insights and their specialized understanding of the business context, the lead technical expert was able to translate this information into a definition of churn that fit SoftCo. With an integrated understanding of the problem, expectations, and required knowledge, the technical experts, together with the head of customer success, defined an action plan that mirrored the technical experts' experience on AI projects, which consisted of three main phases: (1) Data understanding and preparation, (2) Model definition, and (3) Tool integration. Below, we examine how the design choices in each of these phases included different design decisions to deal with the tension between modularity and integration that is typical to AI development. Moreover, we show the organizational consequences of these design decisions, and we unpack what is being mirrored in each phase.

4.4.1 Phase 1: Data Understanding and Preparation

In the first phase, the technical expert set out to construct an elaborate dataset to be used for model development. To better understand the data, the technical expert, together with business operation members and solution architects, started discovering the different data sources identified as relevant. To do this, the lead technical expert asked the business experts to identify and evaluate potentially relevant data sources, to later explore and predict customer churn. As the lead technical expert explained:

> "So, first of all, we had a feedback session. So, I had a meeting with Victor[1] (Head of Customer Success). I said: 'What do you think can impact the probability of churn before building the data?' OK, he said: 'I think product details are informative, so maybe by the number of products you can see'. It was a guess; he was purely saying it based on his business knowledge and business expertise ... They gave me the ideas." – Lead technical expert

> "Based on the history with customers, what are the things we can learn and things that could be a predictor for churn? And then I think, OK, we do communications with them, we have history about their purchases. Uh, yeah, we have all the insights on whether they use e-learning, etcetera. And so, these are all data sources I could think of that we could learn from and could indicate whether a customer was more likely to churn." – Head of Customer Success

One of the first issues that the technical experts encountered was that, the information was collected in different sources, formats, or platforms. For example, the data about contracts and customer information was stored in the CRM platform, while support case information was retrieved from another platform where the cases were created. This modular distribution allowed separate access to specific information by different departments of the organization. However, since the development of the AI tool required an integrated view of customer data, it was essential to aggregate these diverse data sources to enrich the results. Yet, the technical expert quickly found out that, while he put in a lot of work to understand from the business context which data sources could be useful and available for predicting churn, the actual data availability did not match these expectations, and its modular nature presented additional challenges. For instance, when he explored information about the customer usage of the software, inconsistencies with the data available came up, impeding its inclusion in the model:

> "Usage data could not be included because they started to collect it at a different time, then ... you are supposed to have the same timeline (for all the data) if you want to integrate multiple sources ... but even if it would have had the same timeline, not all users were sending their data because it was optional, so you would say 80% of the data is blank if you included usage data just because only 20% shares usage data." – Lead technical expert

Moreover, concerning the data preparation, issues regarding the data quality came up from other data sources. To try and solve the data issues, the lead technical expert studied the data available in the company and selected what he could and could not use. This also required him to adjust the list of predictive variables he initially expected to include. In other words, problems with the data could lead to problems with modelling the AI tool. To further solve the data quality issues, the lead technical expert conducted an iterative data quality verification process with the business experts. This involved working closely with experts from both the operations department and the solution architecture team:

[1]All original names have been removed; the names mentioned are pseudonyms.

> "There was a data quality issue ... In the CRM platform, you're supposed to fill in the seller's name, and the contract, and I was using that variable ... But in some cases, there was a blank value... I worked with John (Operations specialist), as he also knows that data because he worked in order processing before. So, the first year was a kind of exploration ... to understand what happened to data quality and then correcting or fixing. All the first year of understanding and six months of the second year, we spent six months cleaning the data." – Lead technical expert

As such, the integrated design of the dataset required the technical and business experts to 'break the mirror' of the organizational structure represented in the data, allowing the technical expert to tackle the data inconsistencies by leveraging the knowledge and experience of the business experts within the company. This mirror-breaking was an iterative and manual process where the lead technical expert sat down mainly with business operations experts from diverse areas. For instance, in some cases, the field related to the corresponding seller was left empty. To ensure completeness, the lead technical expert collaborated with the account managers so based on their knowledge, they could manually fill in the missing information. In other cases, they had to collectively trace back the renewals of the contracts over time for the cases when there was no information registered.

Over time, they figured that one of the main sources of inconsistencies was the fact that the data was mainly human generated, which made the data susceptible to errors in the data entry process. As the lead technical expert explained: "I asked: Why is this column blank? They said it's blank because we configured it like this or they said: I don't know... then they explained to me that there's human input, that's why it was wrong. So, when I started, I easily noticed the impact of human input."

To solve this issue, the technical expert proposed the solution architect to incorporate additional validations in the process of data collection to prevent human errors securing data quality for future projects. For example, they modified the field of the seller to be mandatory as all the contracts should be assigned to a specific seller, in this sense human errors could be reduced. These changes, although not initially part of the project's scope, ultimately led to the implementation of further developed data collection practices within the company.

After completing the validation and correction of the data, the technical experts created a descriptive analysis of the variables and shared the findings with the head of customer success, head of product, and CEO, to validate whether the data reflected the reality of SoftCo.

> "The data was prepared then we presented descriptive insights from the data to business users to say, (success manager, software engineer, and CEO) ... yeah, mainly these three people. They said that represents our business. So, it was in line with how the business was doing." – Lead technical expert

However, while this phase of breaking the mirror to create an integrated dataset helped to solve issues with the modular data collection practices, one of the consequences was that the expectations of the technical experts regarding the data also changed to the point where they became disillusioned about the richness of the data:*"Victor (Head of Customer Success) told me*

in the beginning: 'We have rich data, we have been collecting it since 2004'. In the end, we ended up using data only since 2017, so all the data from 2004 till 2017 was lost because the setup that they used to collect the data was very difficult to use in our base table." (Lead technical expert).

Overall, the first phase required collaboration among diverse experts across multiple steps, including data understanding, correction, selection, and preparation. Given the modular structure of the data, a thorough analysis was essential for identifying and resolving inconsistencies. To ensure a successful integration of different data sources into a final base table, technical experts needed a comprehensive understanding of the dataset.

4.4.2 Phase 2: Model Definition

Following the creation and validation of the dataset, the technical experts evaluated a variety of potential predictive models. In their decision-making process, they focused not only on predictive performance but also prioritized aspects such as interpretability and explainability, given their importance for the company. They presented and explained the different alternatives to the business experts, to understand their preferences and make decisions together about the best approach to follow. For instance, during one of the meetings, the lead technical expert presented a comparison between so called "black box" and "white box models", denoting the differences in terms of how intuitive the models can be, and outlining some examples of the algorithms for each case.

> "After we presented the descriptive insights, they said yeah, we [business experts] want to see the model, and I did communicate that to [the rest of the technical team] and we proposed diverse options in terms of whether the model ... would be interpretable or not. So, we wanted to show them [business experts] what kind of model could be chosen and what are the advantages and disadvantages, when we presented it to them they decided that they wanted to go for a simple interpretable model." – Lead technical expert

The model that was selected, therefore, mirrored the organizational requirements of simplicity and interpretability. Additionally, to provide more transparency about the decisions taken, the lead technical expert explained the selection process in a clear and easy-to-follow way to the business experts. He used examples and visuals, like charts and diagrams, to make the concepts more accessible for the business experts. The approach he followed, drawing on his technical expertise, began with decreasing the initial set of potential variables to use, removing redundant features, and reducing the complexity of the model. The lead technical expert also explained the process followed for this purpose, and then how to come up with a final set of relevant features to include in the prediction model.

To provide a better understanding of the model selection process, the lead technical expert also explained which evaluation metrics would be used to compare the predictive models, how to calculate them, and their interpretations. Familiar with concepts such as AUC and Lift charts, the business experts were able to accurately interpret the technical outcomes: *"The cumulative gain chart ... instead of randomly reaching 2 out of 10, we can reach 5 out of 10 potential*

churners, so the tool is really good if the evaluation shows that we are pretty good at indicating which customers might churn and which not." (Head of Customer Success).

Next, the technical experts presented the findings of the predictive churn model to the business experts. The explanation focused on key factors influencing customer churn and the methodologies used to reach these results. These interactions and the different validation processes ensured that both types of experts agreed that the final model could meet the requirements in terms of performance and interpretability:

> "That was always an interesting exercise that we did. It's like, OK, what does the business think and what does the data tell us, and we match those two ... There was definitely the link between the business and the data science ... so we used both insights to come up with the model." – Head of Customer Success

In addition, the technical team explained the added value of the tool for the business by showing some examples of the churn prediction calculation for specific cases. They presented the values of the variables for each case and how each variable impacted the predicted probability of churn. In this sense, the technical experts provided clear examples to demonstrate and guarantee the transparency of the developed model. For the company, this transparency entailed not only a comprehensive understanding of the model's functionality and the data it incorporated but also the extent to which the insights derived from the model aligned with their business knowledge and definition of churn.

The dynamics explained, illustrate how organizational requirements were mirrored by the technical experts, which led the final model to evolve from a modular model focused on performance to an integrated model focused on meeting SoftCo's organizational requirements. The mirroring ensured that the final tool met the organizational needs for transparency and interpretability, as business experts believed it would facilitate its integration into the partners' workflow. As the Software architect described: *"It [AI tool] increases awareness, it raises with partners that customers could be at risk of churn... you have to actively pursue renewal of maintenance because if you don't, there's churn."*

4.4.3 Phase 3: Integration of AI tool

Once the organization had a final and explainable churn prediction tool, the goal was to integrate it within the business processes so the channel partners (the final users) could use the tool to support their retention efforts. For this phase, the way in which the company structure incorporates its channel partners is relevant. Different departments engage with these partners depending on the nature of the communication or issue at hand. For instance, to address technical problems, partners connect with the support team or software architects. In contrast, for new business development, they are in touch with the channel managers from the sales team. Concerning the recurrent business —specifically the follow-up of churn rates and retention efforts for existing customers— the direct touchpoint between the organization and the partners is the customer success team. As a customer success manager explained:

> "... the main role of the team [Customer success] is to protect the recurring revenue part, so that is subscriptions and maintenance contract renewals, and we do that also together with the channel partners ... we have a split going on between sales, focusing on new business development, finding new customers and then customer success focusing on the renewals, but together with the channel (partners)." – Customer success manager

Due to the modular communication structure with partners and given the purpose of the churn prediction tool, the customer success team was responsible of presenting and explaining the tool to the channel partners. This task was primarily handled by the head of customer success, who was directly involved in the tool's development. He specifically explained what the churn prediction tool was, how it was developed, and its added value in identifying potential churners: "*I gave a demo on one of the virtual events that we were having, so I presented that, I showed it, I explained the whole reasoning behind it; how we built it, what it was.*" (Head of Customer Success)

While the customer success team expected that their assessment of the transparency and explainability would be enough to meet the expectations of the channel partners, the channel partners quickly started pointing out issues with understanding the churn predictions:

> "Yeah, we have seen it [the tool] because we have access to the CRM portal, but sometimes it's a little bit hard to find out what are the criteria behind, because that's something we don't know, why is the churn rate for a certain customer higher than for another customer? " – Partner 5

Due to these issues, the partners also started to question what the churn predictions generated by the AI system actually reflected. As one of the partners explained:

> "I would like to at least have a question mark on where the criteria come from to give such a high churn rate for a customer or not. Because we have been, or we are in contact with the customer ... which is not the case with [SoftCo]. So, the criteria must come from different angle... Sometimes it's a little bit untransparent." – Partner 5

This "different angle", as expressed in the quote above, conflicted with what the channel partners considered key to their work: the relational nature of the channel-based sales model. As partners maintain a closer relationship with end customers, the partners expressed that their direct communication helps them detect early signs of churn: *"Some relationships with some customers are so old. So, I know beforehand if they're going to renew this or not. So, whatever actions I do, if they have made up their mind that they're not going to renew, they're not going to renew."* (Partner 7)

Moreover, as they operate outside SoftCo, partners play a vital role in gathering customer feedback and insights. Therefore, regarding the reliability of a churn prediction AI tool, they expected these insights to be reflected in the predictions. However, their perception was that the

metrics identified and used by SoftCo fell short in accurately reflecting the reality and nuances of the customers, which they felt were not considered when calculating metrics such as churn rate:

> "Based on what can you say a customer that renewed every year, ... will not renew, I mean renew on time that you can see statistically [sound] but that has nothing to do with what the real number is... a lot of customers, they are late confirming that they will renew the contract. I know that's an issue for [SoftCo], I don't see it for me as a big issue ... if they are one or two weeks earlier, or one or two weeks late, it's the same administration." – Partner 6

Finally, regarding customer retention, partners have differing views from the company on the importance of preventive actions. In some instances, they may prioritize reactive measures based on their understanding of their customers. This aspect impacted the relevance of the churn prediction tool as well and its conception which was based on the preventive approach of the company.

> "So even before the date is passed, for me that is only annoying. I don't want to have any more questions or emailing or anything until at least one month passed. Then you can be worried, but if you are worried before it even passed the date that is not effective. So, what I found out is sending those reminders two months in advance they will be confused and then they will forget. So, I never do that. I think one month in advance is well in time" – Partner 6

In sum, in the design process of the churn prediction AI system, the tool ended up mirroring SoftCo's definition of churn and explainability requirements. However, the challenges encountered during its integration into the partners' workflows highlight differences between the company's perspective and that of the partners, who are traditionally organized in a decentralized, modular way but were considered highly interconnected with the organizational principles of SoftCo in the development of the AI system. This, however, was not an intentional effort of 'breaking the mirror' in the technology design, but an overlooked misalignment that hindered the integration of the tool in the final work process which we call *distorting the mirror*. Figure 4.3 summarizes the empirical process.

4.5 Discussion

Building on the findings of our case, we offer a general overview of how organizations cope with the tension between modularity and integration in AI development, and with what consequences for organizational structure, by showing how mirroring occurs in the case of AI development. By unpacking three phases in AI development – data understanding and preparation, model definition, and tool integration – we identified how technological and organizational requirements triggered a mirroring process. More specifically, in the data understanding and preparation phase, technological requirements for data integration triggered organizational adjustments to overcome its modular design. In the model definition phase, organizational requirements for

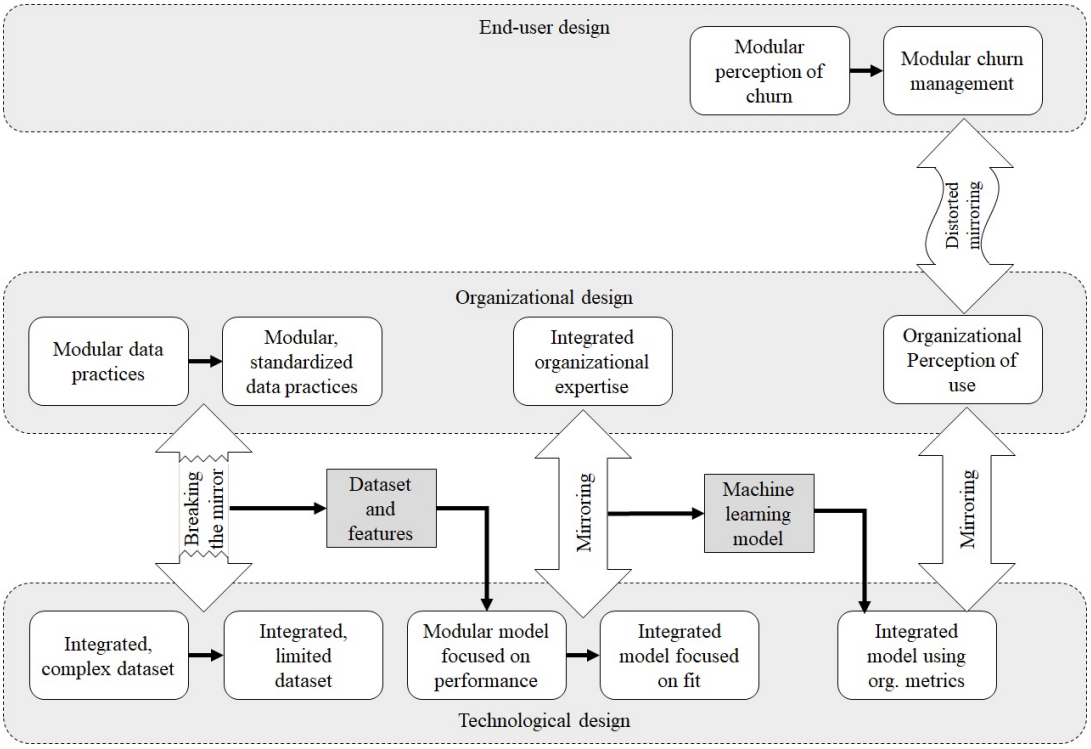


Figure 4.3: Summary empirical process

expertise integration triggered changes to the model to move from a modular design focused on predictive performance to an integrated design focused on organizational fit. Yet, in the final phase, we highlighted how, through this mirroring process, the AI tool also ended up what we call “distorting the mirror", representing the disconnect between the organizational design and the work design of the end user. As such, the model ended up mirroring the host organization’s perception of use, instead of the end users’ requirements, ultimately hampering the usefulness of the tool in practice. Our findings offer contributions to the literature on the mirroring hypothesis and AI development as a co-creation process.

4.5.1 A Processual Understanding of Mirroring

One of the key findings of our study is that different phases of AI development lead to different mirroring outcomes. This offers new insights for the mirroring hypothesis literature.

First, our study aligns with current research emphasizing that AI development is a multistakeholder process (e.g., Bailey and Barley (2020)). By unpacking the role of multiple stakeholders in the development process, our study moves beyond studying clearly defined boundaries (Colfer and Baldwin, 2016) and brings to the fore that understanding mirroring strictly through the lenses of organizational and product design falls short to explain the different dynamics and tensions that influence what is being mirrored and with what consequences for technology and the organizational design. As such, our study shows that, in the context of AI development, the question about “what" is being mirrored is not a straightforward answer which goes beyond unpacking the relationship between modularity and integration in technological and organizational designs.

Second, in contrast to existing literature that mainly views mirroring as a one-time event (Colfer and Baldwin, 2016; Burton and Galvin, 2022), our study highlights that mirroring is accomplished through multiple, cumulative phases, thereby emphasizing the processual nature of mirroring. In this mirroring process, each phase requires close analysis, since different stakeholders can be involved which can influence requirements regarding modularity or integration. For example, as evidenced by the case study, the first phase of data understanding and preparation required an integrated approach. The technical expert discovered that the data sources were collected from different systems; therefore, to create a comprehensive dataset for running machine learning models, he had to collaborate with business experts from multiple areas. This also triggered organizational adjustments to facilitate the integrated dataset. While this example illustrates a case of "mirror breaking", in which the product requirements drive interactions that diverge from the existing organizational structure, this was not the end state of the mirroring process. As such, our study shows that to better understand how mirroring comes about in complex design settings such as AI development requires a more micro-level analysis of the different mirroring practices.

4.5.2 AI Co-creation as a Mirroring Process

Another key finding from our study is that mirroring plays an important role in determining what is inscribed into machine learning algorithms and what is not. This offers insights into the emerging literature on AI development as co-creation.

Current studies on AI development have pointed at the subjective nature of constructing AI systems and the relevance of interactions between diverse experts to make these systems fit the organizational requirements (e.g., Lebovitz et al. (2021); Van den Broek et al. (2021). This perspective relates to the literature on AI development as co-creation, where multiple stakeholders are mutually dependent on each other (Bailey and Barley, 2020; Waardenburg and Huysman, 2022). Our study shows that, as stakeholders co-create, the knowledge exchanged becomes crucial, defining not only the outcome of a particular step but also laying the groundwork for future interactions.

In addition, and according to the above-mentioned insight, it is essential for the co-creation process to define who is involved in each phase and how. We emphasize that, consequently, the AI system will mirror the underlying perspectives and assumptions of the stakeholders involved. For example, our findings show how during the co-creation process, business experts were involved across all three phases and how their knowledge and understanding of churn were ultimately mirrored in the AI tool. However, this also highlighted their subjective understanding of how the tool fits into the workflows of end users, who were not involved in the co-creation process and who ultimately failed to integrate the AI tool in their work. As such, our study shows that, in a co-creation process such as AI development, mirroring can also inflate existing communication barriers or bottlenecks in the organizational design.

4.5.3 Limitations and Future Research

It is crucial to emphasize the limitations of our study, which also provide avenues for future research. Our case represents insights on AI development within a multi-stakeholder setting, a scenario that is increasingly relevant as organizations across different industries are becoming more inclined to adopt AI technologies. However, initial AI implementations often present unique challenges, particularly as they are a first approach into data integration and validation practices. We recognize that an organization's learning curve in AI implementation can potentially influence its strategies for "mirror breaking" and the inclusion of diverse stakeholders. Furthermore, the knowledge gained by technical and business experts from prior AI implementations can affect the dynamics of the mirroring process. Despite these challenges, the findings of this study establish a foundation for exploring the relationship between technological and organizational designs through the lens of the "mirroring hypothesis". This approach contributes to a deeper understanding of the complexities and variations that arise during different phases of the process, motivating future studies to understand subsequent AI implementations, contributing to the understanding of how they evolve in terms of mirroring, but also identifying the relevance of the different designs over time.

Moreover, a relevant path for exploration is the further development of AI tools following their initial implementation, given the evolving nature of these technologies and the adjustments that may occur once they are in use. The potential feedback loops in this process can involve new mirroring dynamics, where internal factors, such as managerial changes, or external factors, such as shifts in the industry, may require the mirroring of additional structures in certain phases. By acknowledging these limitations, we aim to establish a foundation for future studies that can build upon our findings and investigate the complex relationship between technological and organizational factors in AI implementations.

4.5.4 Practical Implications

This study offers practical implications for managers, business experts, and technical experts involved in the development of AI tools in organizations. Considering the complexity of a multi-stakeholder setting for AI development, it is essential for organizations to cope with the tension between modularity and integration in AI development. As evidenced in the case study, organizations can already put in place mechanisms to enable collaboration between diverse stakeholders at specific points during AI development. Nevertheless, this process does not guarantee that the final tool aligns with organizational structures that contribute to its use. Our study emphasizes not only on the relevance of fostering organizational ties between teams to facilitate co-creation, but also understanding which stakeholders should be involved in each phase, as this impacts the continuous development and the structures mirrored by the final AI tool. A clear understanding of the evolving dynamics in the mirroring process, including what should be mirrored at each phase, can facilitate the effective integration of relevant stakeholders throughout each phase. For example, when AI development requires alignment with organizational interpretability standards, it becomes crucial to engage business stakeholders to define the tool's

interpretability level, thereby ensuring their needs are met.

This study also illustrates the complexity of mirroring, emphasizing the need of incorporating end-user design into this process due to the transversal impact of AI tools within organizations. However, in addressing the tension between integration and modularity inherent in AI developments, it is neither feasible nor essential to involve every stakeholder in every phase of the process. By analysing the activities and phases involved in the development process, organizations can detect relevant interaction points between stakeholders and end-users, facilitating communication channels and, if necessary, “breaking the mirror" not only based on technical or organizational designs but also based on user design.

4.6 Conclusion

The interplay between modularity and integration in the development of AI tools within organizations represents an increasingly complex task, increasingly common due to the added value AI brings to organizations. In this study, we examine a case of AI tool development within an organization through the lens of the mirroring hypothesis. Although the mirroring hypothesis has been widely explored in various contexts, the evolving, multi-stakeholder nature of AI technologies underscores the need for dedicated studies that apply this perspective specifically to AI developments. As our findings indicate, the co-creation process of an AI tool development within organizations evolves through interactions and design structures mirrored throughout the process. However, when end-user design is not accurately mirrored, the resulting tool may deviate from the original objectives, leading instead to a ‘distorted’ interpretation of that design. This study of technological and organizational designs in AI development through the mirroring hypothesis highlights the dynamic nature of mirroring in this context and demonstrates how the diverse stakeholders involved in this co-creation process shape the final outcome.

4.7 Appendix

Appendix A. Sample Channel partners Interview Guide

Purpose: Gain a better understanding of the working dynamics of Channel partners Pre-Interview:

- Request consent for recording agreement
- Introduction/purpose of interview
- Request of company’s information (experience, market, types of products you sell, Distributor, reseller etc.)

Initial Questions
Subtopic 1:

Q1) Could you explain the typical process followed by your company when interacting with customers who purchase products from *SoftCo*? Specifically, I would like to know how the first contact is initiated and how you interact with customers after their first purchase.

Q2) Regarding customer retention, could you please describe the specific actions that you or your company take to prevent customer churn?

In case they don't mention it:

- **Q2.1)** Could you please tell me what is your role in terms of customer relationship management?
- **Q2.2)** If I understand correctly, in case of ... you do ...
- **Q2.3)** Which areas / roles are involved in task of customer retention within your company?

Q3) How does the revenue (type of product) of the products influence the actions taken to prevent customer loss?

Q4) If a customer is considered high-value, are different retention strategies implemented compared to those used for low-value customers?

- **Q4.1)** If so: Could you please tell which?

Q5) After taking these actions, which trends or behaviors have you observed in customers? Reasons

Q6) Could you give me some examples?

Subtopic 2: Based on my understanding, *SoftCo* has added a module to Salesforce that enables the viewing of the top 10 customers who are at a higher risk of leaving...

Q7) Is your company familiar with this tool? Additionally, does your company utilize this tool for customer retention purposes?

- **Q7.1)** If so: Could you please tell me how?
- **Q7.2)** If no: Why do you think your company is not using it?

Q8) What factors do you believe this tool should take into account when predicting customer turnover? In other words, what information do you consider to be relevant and should be utilized for these predictions?

Q9) (In case they are familiar with the tool) In your opinion, what improvements can *SoftCo* make to this tool to better assist your company in customer retention tasks?

Deeper Questions Subtopic 3:

Q10) Could you please tell me, what does good customer service mean to you?

Q11) From your experience, what do you believe are the reasons why some customers decide not to renew their maintenance contract with *SoftCo*?

Q12) Besides customer turnover, what other situations do you consider to be relevant in the context of customer success management?

Subtopic 4:

Q13) What do you consider is the role of *SoftCo* in customer retention?

Q14) If I understand correctly, you consider *SoftCo* should be in charge of ...

Q15) Could you please describe the interactions, relationship between your company and *SoftCo*?

Q16) Based on your experience, how do you think *SoftCo* could support your company in improving customer retention?

Wrap up.

Q17) Is there something else you would like to add regarding the topic?

Conclusion of interview

References

Bailey, D. E. and Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, 30(2).

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., and von Krogh, G. (2022). We Are All Theorists of Technology Now: A Relational Perspective on Emerging Technology and Organizing. *Organization Science*, 33(1):1–18.

Baldwin, C. Y. (2023). Design rules: past and future. *Industrial and Corporate Change*, 32(1):11–27.

Baldwin, C. Y. and Clark, K. B. (2000). *Design Rules*, volume 1. The MIT Press.

Burton, N. and Galvin, P. (2022). Modularity, value and exceptions to the mirroring hypothesis. *Journal of Business Research*, 151:635–650.

Cabigiosu, A. and Camuffo, A. (2012). Beyond the "Mirroring" Hypothesis: Product Modularity and Interorganizational Relations in the Air Conditioning Industry. *Organization Science*, 23(3):686–703.

Campagnolo, D. and Camuffo, A. (2010). The concept of modularity in management studies: A literature review.

Chen, G., Huang, L., Xiao, S., Zhang, C., and Zhao, H. (2024). Attending to Customer Attention: A Novel Deep Learning Method for Leveraging Multimodal Online Reviews to Enhance Sales Prediction. *Information Systems Research*, 35(2):829–849.

Colfer, L. and Baldwin, C. Y. (2010). The Mirroring Hypothesis: Theory, Evidence and Exceptions.

Colfer, L. J. and Baldwin, C. Y. (2016). The mirroring hypothesis: theory, evidence, and exceptions. *Industrial and Corporate Change*, 25(5):709–738.

Haenlein, M. and Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4):5–14.

Henderson, R. M. and Clark, K. B. (1990). Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative science quarterly*, 35(1):9–30.

Holland, J. H. (1992). Genetic Algorithms. 267(1):66–73.

Hopfield, J. J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Langlois, R. N. (1992). External Economies and Economic Progress: The Case of the Microcomputer Industry. Technical Report 1.

Lebovitz, S., Levina, N., and Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly: Management Information Systems*, 45(3):1501–1525.

Ma, T., Hu, Y., Lu, Y., and Bhattacharyya, S. (2024). Customer Engagement Prediction on Social Media: A Graph Neural Network Method. *Information Systems Research*.

MacCormack, A., Baldwin, C., and Rusnak, J. (2012). Exploring the duality between product and organizational architectures: A test of the "mirroring" hypothesis. *Research Policy*, 41(8):1309–1324.

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (1955). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955.

Meyer, S., Faria, P., Ganapa, S., Liampoti, A., and Rahne, U. (2023). The BCG X Playbook:Winning strategies of hypergrowth SaaS champions. Technical report.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative Data Analysis: An expanded sourcebook*. SAGE, 2 edition.

Nilsson, N. J. (1980). *Principles of artificial intelligence*. Morgan Kaufmann.

Puranam, P., Raveendran, M., and Knudsen, T. (2012). Organization design: The epistemic interdependence perspective.

Ravasi, D. and Stigliani, I. (2012). Product Design: A Review and Research Agenda for Management Studies. *International Journal of Management Reviews*, 14(4):464–488.

Sanchez, R. and Mahoney, J. T. (1996). Modularity, Flexibility, and Knowledge Management in Product and Organization Design. Technical report.

Shrestha, Y. R., Krishna, V., and von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, 123:588–603.

Sjödin, D., Parida, V., Palmié, M., and Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, 134:574–587.

Sorkun, M. F. and Furlan, A. (2017). Product and Organizational Modularity: A Contingent View of the Mirroring Hypothesis. *European Management Review*, 14(2):205–224.

Sosa, M. E., Eppinger, S. D., and Rowles, C. M. (2004). The misalignment of product architecture and organizational structure in complex product development. *Management Science*, 50(12):1674–1689.

Turing, A. M. (1950). Computing machinery and intelligence.

Van den Broek, E., Sergeeva, A., and Huysman, M. (2021). When the machine meets the expert: An ethnography of developing ai for hiring. *MIS Quarterly: Management Information Systems*, 45(3):1557–1580.

Von Krogh, G. (2018). Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing. *Academy of Management Discoveries*, 4(4):404–409.

Waardenburg, L. and Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, 32(4).

Wang, T., He, C., Jin, F., and Hu, Y. J. (2022). Evaluating the Effectiveness of Marketing Campaigns for Malls Using a Novel Interpretable Machine Learning Model. *Information Systems Research*, 33(2):659–677.

Yin, R. K. (2009). *Case study research design and methods*. SAGE, 4 edition.

Chapter 5

Conclusions

Chapter 5
Conclusions

5.1 Conclusions

5.1.1 Conclusions générales

Cette thèse visait à approfondir la compréhension de la manière dont l'IA est intégrée dans les contextes B2B. Motivée par l'importance croissante de l'IA en tant que moteur de la prise de décision fondée sur les données, la recherche a abordé trois questions interconnectées : comment les applications d'IA créent de la valeur, comment leur intégration est contrainte par des défis analytiques et organisationnels, et comment elles sont mises en œuvre au sein des entreprises. Pour examiner ces dimensions, la thèse a développé un cadre analytique tripartite englobant les Applications, les Défis et l'Implémentation, et a étudié chacune d'elles au travers d'une série de trois études empiriques menées dans un environnement logiciel B2B européen.

À travers ces études, la thèse a montré que l'intégration de l'IA dans les contextes B2B n'est pas un processus purement technique consistant à appliquer des algorithmes à des données, mais un processus organisationnel cumulatif dans lequel la valeur émerge de l'interaction entre les actifs de données, les configurations analytiques et les structures qui régissent la collaboration et la prise de décision. Les résultats illustrent collectivement que la création, la limitation et la réalisation de la valeur fondée sur l'IA sont façonnées par les propriétés spécifiques des contextes B2B — cycles décisionnels prolongés, parties prenantes hétérogènes et structures de coordination multiniveaux — qui les différencient des domaines orientés consommateurs.

Applications de l'IA dans les contextes B2B. La première dimension du cadre concerne la manière dont les technologies d'IA génèrent de la valeur lorsqu'elles sont associées à des sources de données appropriées, des conceptions analytiques adaptées et des finalités organisationnelles claires. Les résultats de cette thèse montrent que, dans les environnements B2B, la création de valeur via l'IA dépend à la fois de la sophistication analytique et de la manière dont données, modèles et acteurs organisationnels sont configurés pour se compléter mutuellement.

Le Chapitre 2 a montré que la transformation de données d'usage de haute dimension en variables prédictives structurées étend la base informationnelle de l'analytique de rétention et fournit une méthode reproductible pour opérationnaliser les données comportementales. Le Chapitre 3 a étendu cette logique analytique à des contextes temporellement dynamiques, en montrant que la modélisation dynamique permet de capturer des schémas d'engagement utilisateur en évolution et de produire des insights précoces et actionnables pour la gestion de l'acquisition. Le Chapitre 4 a ajouté une dimension complémentaire en montrant que la conception et le déploiement de ces systèmes analytiques sont continuellement façonnés par des choix organisationnels et la co-création entre parties prenantes.

Pris ensemble, ces résultats font progresser une compréhension multidimensionnelle des applications de l'IA dans les contextes B2B. Ils montrent qu'une création de valeur efficace par l'IA émerge de l'évolution coordonnée de trois couches interdépendantes : des sources de données qui encodent les interactions

business, des modèles analytiques qui traduisent ces interactions en insights prédictifs, et des mécanismes organisationnels qui convertissent ces insights en action. Les études positionnent ainsi les applications d'IA comme des systèmes adaptatifs qui transforment les données en capacité managériale par un alignement itératif entre sophistication technique, signification analytique et pratique organisationnelle.

Défis analytiques et organisationnels. La deuxième dimension du cadre porte sur les conditions qui limitent le potentiel de l'IA à générer de la valeur en pratique. À travers les études, ces défis se révèlent à la fois analytiques — issus de la complexité des données et des modèles — et organisationnels — découlant des structures de coordination et des exigences d'interprétabilité. Le Chapitre 2 a mis en évidence les pressions techniques et opérationnelles liées à l'intégration de données d'usage à grande échelle dans l'analytique prédictive. Le cadre proposé pour structurer des données comportementales hétérogènes a illustré les arbitrages méthodologiques entre granularité des données, effort de prétraitement et faisabilité computationnelle. L'étude a également élargi la conceptualisation des contraintes analytiques en montrant que l'expérimentation sur de grands ensembles de données comporte des implications en matière de ressources et de durabilité, introduisant l'empreinte environnementale de l'IA comme préoccupation managériale émergente.

Le Chapitre 3 a approfondi cette perspective en révélant une seconde couche de défi : la tension entre sophistication algorithmique et interprétabilité managériale. Les architectures d'apprentissage profond ont amélioré la précision prédictive, mais au risque d'opacité, fragilisant la confiance organisationnelle et la pertinence pour la décision. En traduisant des sorties de modèles complexes en trajectoires comportementales interprétables, l'étude a montré que répondre à ces défis suppose de concilier progrès technique et clarté communicationnelle, afin de garantir que les insights analytiques demeurent utilisables dans les environnements managériaux.

Le Chapitre 4 a étendu l'analyse au domaine organisationnel, montrant que les tensions de conception entre modularité et intégration se manifestent sous forme de désalignements entre systèmes technologiques et structures organisationnelles. Ces frictions sont d'ordre organisationnel, dans la mesure où elles inscrivent des schémas de communication et des asymétries de pouvoir dans la technologie elle-même. Les résultats soulignent que les défis analytiques et organisationnels sont co-constitutifs : la faisabilité de l'intégration de l'IA dépend simultanément de la stabilité des infrastructures de données, de l'interprétabilité des modèles et de l'adaptabilité des mécanismes de coordination.

Collectivement, ces éclairages font progresser une compréhension systémique des défis d'intégration de l'IA dans les contextes B2B. Ils montrent que les frictions techniques, analytiques et organisationnelles se renforcent mutuellement, créant une contrainte multiniveau qui limite la scalabilité et l'usage pérenne. Surmonter ces barrières requiert non seulement un raffinement algorithmique, mais aussi des processus d'apprentissage organisationnel qui développent des capacités d'interprétation, un alignement interfonctionnel et une sensibilité éthique autour de l'expérimentation et du déploiement de l'IA.

Implémentation et usage pérenne. La troisième dimension se concentre sur la manière dont les systèmes d'IA sont ancrés et maintenus dans la pratique organisationnelle. La thèse montre que l'implémentation n'est pas un point final discret du développement de modèles, mais un processus continu d'alignement entre parties prenantes, flux de travail et mécanismes de gouvernance. Le Chapitre 4 a offert une vue approfondie de ce processus à travers une étude de cas longitudinale du développement et du déploiement d'un système d'IA. L'analyse a montré que l'implémentation se déroule par phases itératives de compréhension des données, de définition du modèle et d'intégration de l'outil, chacune nécessitant des négociations entre experts techniques, responsables métiers et utilisateurs finaux. En mobilisant l'hypothèse de mirroring, l'étude a montré que les systèmes d'IA reproduisent des éléments du contexte organisationnel dans lequel ils sont créés — renforçant parfois une coordination productive, mais amplifiant aussi, dans d'autres cas, les décalages entre hypothèses de conception et pratiques réelles. Une implémentation pérenne requiert donc des mécanismes de rétroaction, d'inclusion des utilisateurs et de réajustement institutionnel pour

prévenir de telles distorsions.

Les Chapitres 2 et 3 ont illustré des formes complémentaires d'implémentation au niveau opérationnel. En intégrant des indicateurs de performance tels que le profit maximum attendu (EMPB) dans la modélisation prédictive, le Chapitre 2 a relié les sorties analytiques aux objectifs financiers, permettant aux modèles de rétention de fonctionner comme outils d'aide à la décision au sein des routines business. Le Chapitre 3 a étendu cette intégration opérationnelle à l'acquisition de clients, en montrant comment des signaux prédictifs temporellement variables peuvent guider un outreach dynamique et l'allocation des ressources. Dans les deux études, les résultats ont montré que l'ancrage de l'IA dans les processus managériaux transforme celle-ci d'un exercice analytique en une capacité continue qui informe, coordonne et s'adapte aux rythmes décisionnels organisationnels.

Pris ensemble, les trois études redéfinissent l'implémentation de l'IA dans les contextes B2B comme un processus d'alignement réciproque — entre insights fondés sur les données et jugement humain, conception technique et routines organisationnelles, performance analytique et pertinence business. La création de valeur durable dépend ainsi de la capacité à cultiver cet équilibre dynamique, dans lequel les systèmes prédictifs évoluent en parallèle des organisations qui les utilisent, garantissant que l'IA demeure à la fois techniquement viable et institutionnellement significative dans le temps.

Création, préservation et institutionnalisation de la valeur de l'IA tout au long du cycle de vie

En synthétisant les trois dimensions du cadre, cette thèse propose une compréhension multiniveau et processuelle de la manière dont l'IA contribue à la valeur économique dans les contextes B2B. Les analyses montrent que l'impact de l'IA dépasse les applications isolées ou les gains de performance à court terme : elle reconfigure la façon dont les entreprises créent, maintiennent et organisent la valeur à travers des processus décisionnels fondés sur les données. Dans cette perspective cumulative, la valeur n'est pas un résultat fixe, mais une propriété évolutive de l'interaction entre données, algorithmes et systèmes organisationnels. En s'appuyant sur les résultats empiriques des trois chapitres, la thèse identifie trois mécanismes interreliés par lesquels la valeur générée par l'IA se déploie — création de valeur, préservation de la valeur et institutionnalisation de la valeur — chacun correspondant à une phase distincte dans l'évolution de l'intégration de l'IA. Ensemble, ces mécanismes révèlent l'IA comme un processus récursif à travers lequel des éléments analytiques, opérationnels et organisationnels co-évoluent dans le temps.

La création de valeur se manifeste au stade de l'acquisition de clients, où l'IA renforce la capacité des entreprises à identifier, qualifier et convertir des prospects en modélisant les dynamiques comportementales qui se déploient lors des premières interactions et des périodes d'essai (Chapitre 3). Sur le plan méthodologique, cela implique d'intégrer l'usage dans la modélisation prédictive via des variables temporellement variables et des évaluations à horizon glissant, permettant aux entreprises d'agir sur des signaux d'engagement émergents plutôt que sur des agrégats statiques. Sur le plan conceptuel, la création de valeur renvoie à la réactivité à court terme de l'IA et à sa capacité à apprendre à partir de données clients en évolution rapide et à transformer des schémas temporels en opportunités actionnables.

La préservation de la valeur apparaît dans les processus de rétention client, où des modèles prédictifs détectent le risque de churn et soutiennent des interventions ciblées et opportunes qui consolident les relations en cours (Chapitre 2). Dans ce contexte, l'IA crée de la valeur sur un horizon de moyen terme : en structurant les données d'usage selon différentes dimensions, elle capture les rythmes récurrents de l'engagement client et les traduit en actions managériales préventives. La préservation reflète ainsi la capacité de l'IA à prolonger et stabiliser la valeur au fil de cycles d'usage répétés, en maintenant la continuité des relations clients.

L'institutionnalisation de la valeur se joue au niveau organisationnel, lorsque l'IA devient intégrée

dans les structures de gouvernance, les infrastructures de données et les routines de coordination qui soutiennent les capacités analytiques dans le temps (Chapitre 4). L'analyse longitudinale montre que l'institutionnalisation est un processus de long terme par lequel les organisations absorbent, adaptent et affinent les pratiques d'IA. Elle implique de développer une compréhension de ce qui doit être aligné entre architectures techniques et structures organisationnelles, et de ce qui doit rester distinct, transformant l'IA d'une initiative ponctuelle de projet en une capacité de co-création durable.

Situés dans le cadre plus large du cycle de vie marketing, ces trois mécanismes enrichissent les modèles classiques d'acquisition et de rétention de clients en y ajoutant une couche organisationnelle à travers laquelle l'IA les relie et les renforce. Plutôt que de constituer des étapes séparées, la création, la préservation et l'institutionnalisation de la valeur permises par l'IA représentent un processus continu qui relie l'analytique client à l'apprentissage organisationnel. Ils situent également les études empiriques dans une architecture temporelle articulant l'amont de la création de valeur, la maintenance relationnelle de la préservation de la valeur et l'ancrage organisationnel à travers lequel la valeur devient institutionnalisée. Collectivement, ces éclairages montrent que la contribution de l'IA à la performance économique dans les contextes B2B dépend de la capacité des entreprises à synchroniser réactivité analytique à court terme, gestion relationnelle à moyen terme et renouvellement capacitaire à long terme. D'un point de vue managérial, réaliser ce potentiel suppose d'aligner l'innovation technologique avec les rythmes organisationnels d'apprentissage et d'adaptation, afin que les systèmes prédictifs évoluent avec les marchés, les relations et les environnements organisationnels.

Pris ensemble, les résultats de cette thèse apportent une réponse aux trois questions de recherche initiales. Premièrement, les études montrent que la valeur prédictive dans les contextes B2B émerge lorsque des données comportementales complexes sont transformées en représentations qui reflètent la manière dont les décisions se déploient au sein d'échanges relationnels de longue durée. Deuxièmement, la recherche démontre que les contraintes analytiques et organisationnelles sont étroitement interconnectées : la performance des modèles dépend des infrastructures de données, des exigences d'interprétabilité et des mécanismes de coordination qui déterminent la manière dont les prédictions sont comprises et utilisées. Troisièmement, les analyses longitudinales illustrent que l'implémentation est un processus itératif, au cours duquel les systèmes prédictifs s'ancrent durablement lorsque les conceptions techniques, les structures de gouvernance et les pratiques des utilisateurs évoluent conjointement dans le temps.

De manière générale, la thèse montre que l'intégration de systèmes prédictifs dans les contextes B2B repose sur l'alignement entre la structuration des données, les approches de modélisation et les processus organisationnels. À travers les études empiriques, la recherche met en évidence que les modèles prédictifs acquièrent une pertinence opérationnelle lorsque leurs fondations de données reflètent les dynamiques propres aux interactions B2B et lorsque leurs résultats peuvent être interprétés de manière significative et intégrés dans les routines quotidiennes. Cette perspective intégrée constitue la contribution centrale de la thèse, en offrant une compréhension plus complète de la manière dont la prédiction fondée sur les données peut soutenir la création de valeur, la continuité des relations et l'apprentissage organisationnel dans des environnements B2B complexes.

5.1.2 Limites et perspectives de recherche

Bien que cette thèse fasse progresser de manière substantielle la compréhension de l'intégration de l'IA dans les contextes B2B, plusieurs limites ouvrent également des perspectives prometteuses pour de futures recherches.

Premièrement, les études quantitatives des Chapitres 2 et 3 reposent sur des données provenant d'entreprises uniques au sein de l'industrie européenne du logiciel. Bien que ces cas fournissent des preuves détaillées et représentatives des environnements SaaS B2B, de futurs travaux pourraient tester la

généralisabilité des cadres proposés dans d'autres secteurs et contextes culturels. Des études comparatives pourraient examiner si la structuration des données d'usage et les dynamiques d'engagement temporel présentent une valeur prédictive similaire dans des industries telles que la fabrication, la logistique ou les services professionnels.

Deuxièmement, les coûts environnementaux et computationnels du développement de l'IA ont été abordés dans cette thèse à un niveau conceptuel à travers l'estimation de l'empreinte carbone. De futurs travaux pourraient prolonger cette ligne d'enquête en intégrant systématiquement des métriques de durabilité dans les cadres d'évaluation des modèles et en explorant des stratégies d'optimisation conciliant précision prédictive, impact business et efficacité écologique.

Troisièmement, bien que les modèles d'apprentissage profond du Chapitre 3 et l'analyse de co-création du Chapitre 4 soulignent l'interprétabilité comme préoccupation centrale, des recherches supplémentaires sont nécessaires pour faire progresser des méthodes d'IA explicable adaptées aux processus décisionnels B2B. De telles méthodes devraient relier les explications techniques au raisonnement managérial, permettant aux organisations de calibrer confiance, responsabilité et autonomie dans les décisions soutenues par l'IA.

Quatrièmement, l'étude longitudinale du Chapitre 4 s'est concentrée sur un cycle unique de développement d'IA au sein d'une seule entreprise. De futures recherches pourraient étudier les itérations suivantes du même système ou des développements comparables dans d'autres organisations, afin de saisir comment les dynamiques de mirroring évoluent à mesure que la maturité de l'IA augmente. En particulier, l'examen de l'adaptation post-déploiement, des boucles de rétroaction et de la diffusion inter-entreprises offrirait des éclairages précieux sur la manière dont les outils d'IA passent de projets pilotes à des capacités organisationnelles intégrées.

En résumé, cette thèse propose une vision intégrée de l'IA dans les contextes B2B en examinant ses applications, ses défis et son implémentation comme des processus interconnectés. Elle démontre que la réalisation du potentiel de l'IA dépend de l'alignement entre innovation méthodologique, pratique organisationnelle et considérations de durabilité. En combinant expérimentation fondée sur les modèles et analyse organisationnelle, la thèse contribue tant à la compréhension scientifique qu'à la pratique managériale, offrant une base pour de futures recherches sur la manière dont les technologies d'IA peuvent être conçues, déployées et implémentées dans des environnements business complexes et multi-acteurs.

5.2 Conclusions

5.2.1 General conclusions

This dissertation set out to advance understanding of how AI is integrated within B2B contexts. Motivated by the growing prominence of AI as a driver of data-driven decision-making, the research addressed three interrelated questions: how AI applications create value, how their integration is constrained by analytical and organizational challenges, and how they are implemented within firms. To investigate these dimensions, the dissertation developed a threefold analytical framework encompassing Applications, Challenges, and Implementation, and examined each through a sequence of three empirical studies conducted in a European B2B software environment.

Across these studies, the dissertation has demonstrated that AI integration in B2B settings is not a purely technical process of applying algorithms to data, but a cumulative organizational process in which value emerges from the interaction between data assets, analytical configurations, and the structures that govern collaboration and decision-making. The findings collectively illustrate that the creation, constraint, and realization of AI-based value are shaped by the specific properties of B2B contexts, extended decision cycles, heterogeneous stakeholders, and multi-layered coordination structures, that differentiate them

from consumer-oriented domains.

AI Applications in B2B Contexts. The first dimension of the framework concerns how AI technologies generate value when matched with appropriate data sources, analytical designs, and organizational purposes. The findings of this dissertation reveal that in B2B environments, the creation of value through AI is a matter of analytical sophistication and how data, models, and organizational actors are configured to complement one another.

Chapter 2 demonstrated that transforming high-dimensional usage data into structured predictive features extends the informational base of retention analytics and provides a replicable method for operationalizing behavioral data. Chapter 3 expanded this analytical logic to time-varying settings, showing that dynamic modeling can capture evolving user engagement patterns and deliver early, actionable insights for acquisition management. Chapter 4 added a complementary dimension by evidencing that the design and deployment of these analytical systems are continuously shaped by organizational choices and stakeholder co-creation.

Collectively, these findings advance a multi-dimensional understanding of AI applications in B2B contexts. They reveal that effective AI value creation emerges from the coordinated evolution of three interdependent layers: data sources that encode business interactions, analytical models that translate these interactions into predictive insight, and organizational mechanisms that convert insight into action. The studies thus position AI applications as adaptive systems that transform data into managerial capability through iterative alignment between technical sophistication, analytical meaning, and organizational practice.

Analytical and Organizational Challenges. The second dimension of the framework, addresses the conditions that constrain AI's potential to generate value in practice. Across the studies, these challenges were shown to be both analytical, arising from data and model complexity, and organizational, stemming from coordination structures and interpretability requirements. Chapter 2 exposed the technical and operational pressures of integrating large-scale usage data into predictive analytics. The proposed framework for structuring heterogeneous behavioral data depicted the methodological trade-offs between data granularity, preprocessing effort, and computational feasibility. The study also broadened the conceptualization of analytical constraints by demonstrating that experimentation with large data sets carries resource and sustainability implications, introducing the environmental footprint of AI as an emerging managerial concern.

Chapter 3 expanded this view by revealing a second layer of challenge: the tension between algorithmic sophistication and managerial interpretability. Deep-learning architectures improved predictive precision but risked opacity, undermining organizational trust and decision relevance. By translating complex model outputs into interpretable behavioral trajectories, the study demonstrated that addressing these challenges requires balancing technical advancement with communicative clarity, ensuring that analytical insights remain usable within managerial settings.

Chapter 4 extended the analysis to the organizational domain, showing that the design tensions of modularity and integration manifest as misalignments between technological systems and organizational structures. These frictions are organizational, as they embed communication patterns and power asymmetries into the technology itself. The findings underscore that analytical and organizational challenges are co-constitutive: the feasibility of AI integration depends simultaneously on the stability of data infrastructures, the interpretability of models, and the adaptability of coordination mechanisms.

Collectively, these insights advance a systemic understanding of AI integration challenges in B2B settings. They highlight that technical, analytical, and organizational frictions reinforce one another, creating a multi-level constraint that limits scalability and sustained use. Overcoming such barriers requires not only algorithmic refinement but also organizational learning processes that build interpretive capacity, cross-functional alignment, and ethical awareness around AI experimentation and deployment.

Implementation and Sustained Use. The third dimension, focuses on how AI systems are embedded and maintained within organizational practice. The dissertation demonstrates that implementation is not a discrete end-point of model development but a continuous process of alignment among stakeholders, workflows, and governance mechanisms. Chapter 4 offered an in-depth view of this process through a longitudinal case study of AI system development and deployment. The analysis revealed that implementation unfolds through iterative phases of data understanding, model definition, and tool integration, each requiring negotiation among technical experts, business managers, and end users. Applying the mirroring hypothesis, the study showed that AI systems reproduce elements of the organizational context in which they are created—sometimes reinforcing productive coordination, but at other times amplifying disconnects between design assumptions and real-world practices. Sustained implementation therefore requires mechanisms for feedback, user inclusion, and institutional recalibration to prevent such distortions.

Chapters 2 and 3 illustrated complementary forms of implementation at the operational level. By integrating performance metrics such as the Expected Maximum Profit (EMPB) into predictive modeling, Chapter 2 linked analytical outputs to financial objectives, enabling retention models to function as decision-support tools within business routines. Chapter 3 extended this operational integration to customer acquisition, demonstrating how time-varying predictive signals can guide dynamic outreach and resource allocation. Across both studies, the findings showed that embedding AI into managerial processes transforms it from an analytical exercise into a continuous capability that informs, coordinates, and adapts to organizational decision rhythms.

Taken together, the three studies reframe AI implementation in B2B settings as a process of reciprocal alignment—between data-driven insights and human judgment, technical design and organizational routines, analytical performance and business relevance. Sustained value creation thus depends on cultivating this dynamic balance, where predictive systems evolve alongside the organizations that use them, ensuring that AI remains both technically viable and institutionally meaningful over time.

AI value creation, preservation, and institutionalization across the lifecycle

Synthesizing across the three dimensions of the framework, this dissertation advances a multi-level and processual understanding of how AI contributes to business value in B2B contexts. The analyses show that AI's impact extends beyond isolated applications or short-term performance gains: it reshapes how firms create, sustain, and organize value through data-driven decision processes. In this cumulative perspective, value is not a fixed outcome but an evolving property of the interaction between data, algorithms, and organizational systems. Building on the empirical findings of the three chapters, the dissertation identifies three interrelated mechanisms through which AI-generated value unfolds, value creation, value preservation, and value institutionalization, each corresponding to a distinct phase in the evolution of AI integration. Together, these mechanisms reveal AI as a recursive process through which analytical, operational, and organizational elements co-evolve over time.

Value creation is evidenced at the customer-acquisition stage, where AI enhances firms' ability to identify, qualify, and convert prospects by modeling the behavioral dynamics that unfold during early interactions and trial periods (Chapter 3). Methodologically, this involves embedding usage into predictive modeling through time-varying features and rolling-horizon evaluations, enabling firms to act on emerging engagement signals rather than static aggregates. Conceptually, value creation captures AI's short-term responsiveness and its capacity to learn from fast-changing customer data and transform temporal patterns into actionable opportunities.

Value preservation emerges during customer-retention processes, where predictive models detect churn risk and support timely, targeted interventions that sustain ongoing relationships (Chapter 2). In this context, AI creates value over a medium-term horizon: by structuring usage data along different

dimensions, AI captures the recurring rhythms of customer engagement and translates them into preventive managerial actions. Preservation thus reflects AI's ability to extend and stabilize value across repeated usage cycles, maintaining continuity in customer relationships.

Value institutionalization takes place at the organizational level, where AI becomes embedded in governance structures, data infrastructures, and coordination routines that sustain analytical capabilities over time (Chapter 4). The longitudinal analysis shows that institutionalization is a long-horizon process through which organizations absorb, adapt, and refine AI practices. It involves developing an understanding of where technical architectures and organizational structures should align and where they should remain distinct, turning AI from a project-based initiative into a sustained co-creation capability.

Framed within the broader marketing lifecycle, these three mechanisms enrich classical models of customer acquisition and retention by introducing an organizational layer through which AI connects and reinforces them. Rather than constituting separate stages, AI-enabled value creation, preservation, and institutionalization represent an ongoing process that links customer analytics with organizational learning. They also situate the empirical studies within a temporal architecture of front-end of value creation, relational maintenance of value preservation, and the organizational embedding through which value becomes institutionalized. Collectively, these insights show that AI's contribution to business performance in B2B contexts depends on how effectively firms synchronize short-term analytical responsiveness, medium-term relationship management, and long-term capability renewal. From a managerial perspective, realizing this potential requires aligning technological innovation with organizational rhythms of learning and adaptation so that predictive systems evolve with changing markets, relationships, and organizational environments.

Taken together, the findings of this dissertation provide an answer to the three initial research questions. First, the studies show that predictive value in B2B contexts emerges when complex behavioral data are transformed into representations that reflect how decisions unfold in long, relationship-based exchanges. Second, the research demonstrates that analytical and organizational constraints are tightly interconnected: model performance depends on data infrastructures, interpretability requirements, and coordination routines that determine how predictions are understood and used. Third, the longitudinal evidence illustrates that implementation is an iterative, process in which predictive systems become embedded when technical designs, governance structures, and user practices evolve together over time.

Overall, the dissertation demonstrates that the integration of predictive systems in B2B contexts depends on the alignment of data structuring, modelling approaches, and organizational processes. Across the empirical studies, the research shows that predictive models achieve practical relevance when their data foundations reflect the dynamics of B2B interactions and when their outputs can be meaningfully interpreted and incorporated into everyday routines. This integrated perspective constitutes the core contribution of the thesis, offering a more comprehensive understanding of how data-driven prediction can support value creation, relationship continuity, and organizational learning in complex B2B environments.

5.2.2 Limitations and future research

While this dissertation advances a comprehensive understanding of AI integration in B2B contexts, several limitations also indicate promising avenues for future research.

First, the quantitative studies in Chapters 2 and 3 are based on data from single-firm settings within the European software industry. Although these cases provide detailed and representative evidence for B2B SaaS environments, future research could test the generalizability of the proposed frameworks across different sectors and cultural contexts. Comparative studies could examine whether the structuring of usage data and the dynamics of temporal engagement hold similar predictive value in industries such as manufacturing, logistics, or professional services.

Second, the environmental and computational costs of AI development were addressed in this dissertation at a conceptual level through carbon footprint estimation. Future work could extend this line of inquiry by systematically integrating sustainability metrics into model evaluation frameworks and exploring optimization strategies that balance predictive accuracy, business impact, and ecological efficiency.

Third, while the deep learning models in Chapter 3 and the co-creation analysis in Chapter 4 highlight interpretability as a central concern, further research is needed to advance explainable AI methods tailored to B2B decision processes. Such methods should bridge technical explanations with managerial reasoning, enabling organizations to calibrate trust, accountability, and autonomy in AI-supported decisions.

Fourth, the longitudinal study in Chapter 4 focused on a single AI development cycle within one firm. Future research could investigate subsequent iterations of the same system or comparable developments in other organizations to capture how mirroring dynamics evolve as AI maturity increases. In particular, examining post-deployment adaptation, feedback loops, and cross-firm diffusion would offer valuable insights into how AI tools transition from pilot projects to embedded organizational capabilities.

In summary, this dissertation offers an integrated view of AI in B2B contexts by examining its applications, challenges, and implementation as interconnected processes. It demonstrates that the realization of AI's potential depends on aligning methodological innovation with organizational practice and sustainability considerations. By combining model-based experimentation with organizational analysis, the dissertation contributes to both scholarly understanding and managerial practice, providing a foundation for future research on how AI technologies can be designed, deployed, and implemented in complex, multi-actor business environments.

List of Figures

List of Tables