Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
Doi: 10.63125/9b316w70

Article

ADVANCEMENTS IN MACHINE LEARNING FOR CUSTOMER RETENTION: A SYSTEMATIC LITERATURE REVIEW OF PREDICTIVE MODELS AND CHURN ANALYSIS

Md Nur Hasan Mamun[1];

[1] *Master of Business Analytics, Trine University, Michigan, USA*
Email: nurmamun1112@gmail.com ; mmamun24@my.trine.edu

Citation:
Mamun, M. N. H. (2025). *Advancements in machine learning for customer retention: A systematic literature review of predictive models and churn analysis. Journal of Sustainable Development and Policy, 1*(1), 250–284. https://doi.org/10.63125/9b316w70

Received:
April 20, 2025

Revised:
May 27, 2025

Accepted:
June 25, 2025

Published:
July 09, 2025

ABSTRACT
Customer retention has emerged as a critical strategic objective for organizations seeking to sustain profitability and competitive advantage, particularly in highly saturated and dynamic markets. Predictive modeling, driven by machine learning (ML) techniques, plays an increasingly essential role in enabling firms to identify customers at high risk of churn and to implement proactive retention interventions. This systematic literature review provides a comprehensive synthesis of contemporary advancements in ML-based customer retention analytics, focusing on predictive models and churn analysis across diverse industries, including telecommunications, banking, e-commerce, and subscription-based services. Utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, 112 peer-reviewed studies published between 2015 and 2025 were rigorously selected to ensure methodological rigor and relevance. The review systematically categorizes ML techniques into supervised, unsupervised, and hybrid approaches, with a particular emphasis on widely adopted algorithms such as logistic regression, decision trees, support vector machines, random forests, gradient boosting machines, and deep neural networks, including convolutional and recurrent architectures. In addition, it critically evaluates feature engineering methods, data preprocessing practices, dataset properties, and model evaluation metrics, including accuracy, precision, recall, F1-score, AUC-ROC, and cost-sensitive measures. Special attention is given to emerging research domains, including explainable artificial intelligence (XAI), real-time predictive analytics, transfer learning, and federated learning, which enhance model transparency, adaptability, and privacy compliance. Findings from the review reveal that ensemble methods and deep learning models consistently outperform traditional classifiers in detecting intricate churn patterns, particularly when behavioral, transactional, and sentiment-based features are integrated. The review also underscores the growing importance of interpretable models, post-hoc explanation techniques such as SHAP and LIME, and privacy-preserving methodologies to address challenges related to algorithmic opacity, ethical compliance, and deployment scalability. By mapping the evolution of ML-driven churn analytics, identifying persistent research gaps, and proposing actionable directions for future inquiry, this study contributes to both the academic literature and practical applications in customer retention strategy development.

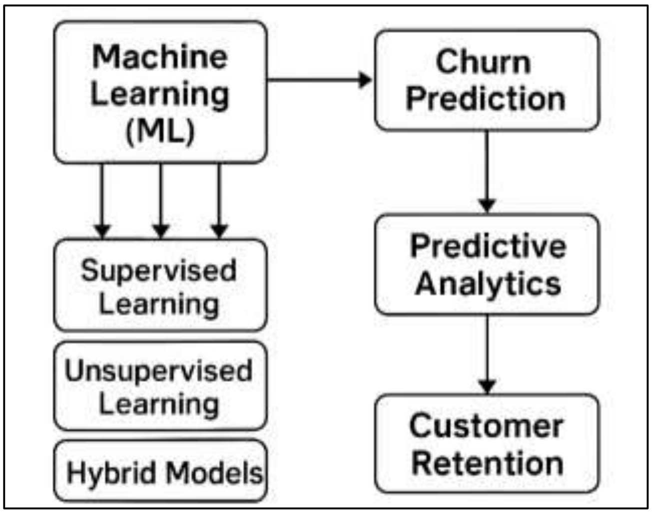
KEYWORDS
Customer Retention; Churn Prediction; Machine Learning; Predictive Analytics; Customer Behavior Modeling;

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

INTRODUCTION

Customer retention refers to a business's ability to keep its customers over a specific period, and it is a key performance indicator reflecting customer satisfaction, product quality, and service effectiveness (Zdziebko et al., 2024). It involves a firm's efforts to encourage repeat purchases and prevent customer churn—defined as the rate at which customers stop doing business with a company (Darzi & Bhat, 2018). Retaining existing customers is substantially more cost-effective than acquiring new ones, with estimates suggesting that acquiring a new customer can cost five to seven times more than retaining an existing one (Sjarif et al., 2020). Therefore, organizations across various industries including finance, telecommunications, and e-commerce have prioritized retention strategies to sustain profitability and maintain competitive advantages. The economic implications are significant globally, as customer churn leads to billions in revenue loss annually, especially in sectors reliant on subscription models or long-term service agreements. The strategic emphasis on retention reflects a shift from transactional marketing to relationship marketing, where the focus is on lifetime customer value and long-term engagement (Prasad & Madhavi, 2012). As markets mature and customer acquisition becomes more saturated and costly, firms must identify patterns of attrition through analytical methods, underscoring the need for accurate and timely churn prediction tools.

Figure 1: Theoretical Framework for Machine Learning-Based Customer Retentio



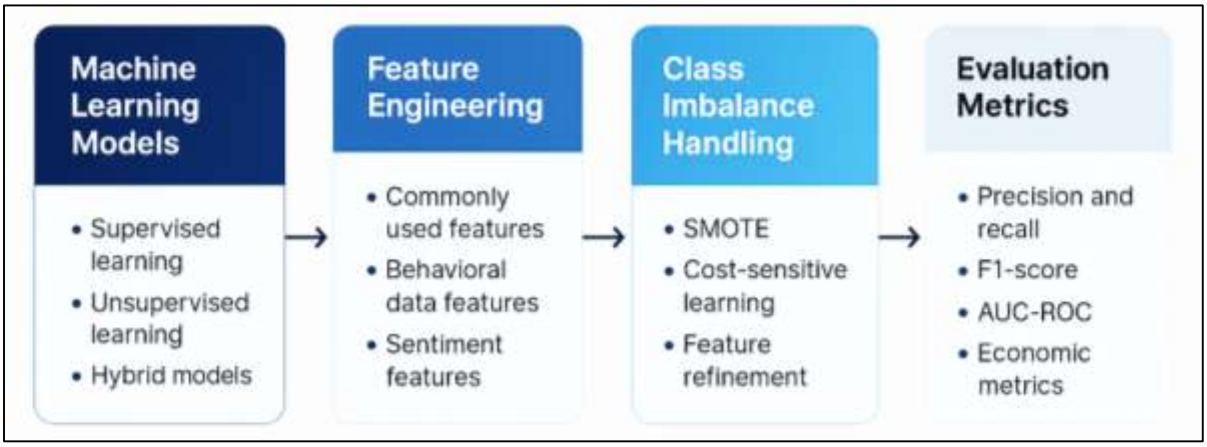
Predictive analytics plays a crucial role in customer retention by enabling firms to proactively identify customers at risk of churning and intervene with tailored retention strategies. It involves the application of statistical techniques and machine learning algorithms to historical customer data to forecast future behavior (Gattermann-Itschert & Thonemann, 2022). The rise of big data has empowered businesses with vast troves of behavioral, transactional, and demographic data, which can be harnessed to develop predictive models with high accuracy (Zdziebko et al., 2024). Early models primarily relied on logistic regression, decision trees, and clustering techniques, but their limitations in capturing nonlinear relationships have spurred a shift toward more sophisticated machine learning approaches. The application of predictive analytics has been particularly prominent in telecommunications, where customer attrition is a critical concern due to market saturation and low switching costs. Banking, insurance, retail, and digital services have also adopted churn prediction frameworks to mitigate loss of high-value clients. Effective churn prediction models not only identify at-risk customers but also help organizations allocate marketing resources efficiently, increasing return on investment. The integration of time-series behavior, real-time feedback loops, and sentiment analysis further enriches the predictive process, enhancing responsiveness to emerging patterns.

Machine learning (ML), a subset of artificial intelligence, has revolutionized predictive analytics in customer retention by offering robust models capable of learning complex patterns without being explicitly programmed (Darzi & Bhat, 2018). Unlike traditional statistical models, ML techniques adapt to new data and improve performance over time, enabling dynamic churn analysis that accounts for evolving customer behavior. Supervised learning methods—such as support vector machines

Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

(SVM), decision trees, random forests, and neural networks—are particularly well-suited for classification problems like churn prediction. Unsupervised learning, including clustering and association rule mining, facilitates customer segmentation and the discovery of latent behavioral patterns. Hybrid models combining both paradigms enhance model generalization and robustness. For example, ensemble learning approaches like gradient boosting and bagging have consistently demonstrated superior accuracy compared to standalone classifiers. Deep learning architectures, including convolutional and recurrent neural networks, further extend the analytical capacity by processing unstructured data such as text reviews and social media activity (Höppner et al., 2020). These advancements allow for more precise and granular predictions, enabling firms to develop targeted and context-sensitive retention strategies. Moreover, ML facilitates automated feature selection, reducing human bias and improving the efficiency of model development. As ML adoption accelerates across industries, its role in optimizing customer lifecycle management continues to expand in significance (Lamrhari et al., 2022).

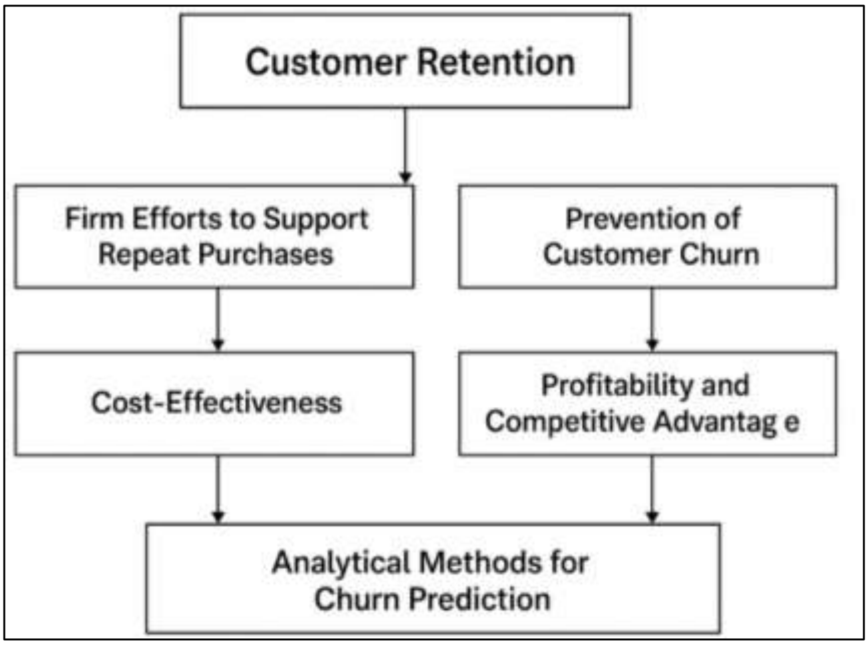
Figure 2: Framework for Machine Learning-Driven Customer Churn Prediction Process



The efficacy of ML models for customer retention hinges not only on the algorithm itself but also on the quality of data and the relevance of input features (Jajam et al., 2023). Feature engineering involves the process of transforming raw data into informative variables that enhance the predictive power of models (Singh et al., 2023). In churn prediction, commonly used features include recency, frequency, and monetary value (RFM), contract type, payment history, customer service interactions, and usage intensity. Behavioral data derived from digital footprints, such as website clicks, app usage, and purchase patterns, are increasingly integrated into feature sets to capture nuanced user behavior. Sentiment features extracted from textual data—such as support tickets, reviews, and social media comments—add a qualitative dimension to churn analysis. The availability of open-source datasets like the IBM Telco Churn Dataset, Kaggle repositories, and proprietary organizational databases has accelerated experimentation and benchmarking in academic and industrial settings. Moreover, advanced data preprocessing techniques including normalization, imputation, and dimensionality reduction (e.g., PCA, t-SNE) are employed to refine datasets and optimize model performance. Class imbalance remains a critical challenge in churn prediction, as churners often represent a small fraction of the customer base. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and cost-sensitive learning are applied to address this skew and ensure balanced model learning. The iterative refinement of features is thus central to the predictive accuracy and interpretability of churn models.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Figure 3: Machine Learning-Driven Customer Retention and Churn Prediction



Assessing the performance of machine learning models in customer churn prediction requires the use of robust and context-sensitive evaluation metrics. Standard metrics include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). While accuracy provides a basic measure of correctness, it can be misleading in imbalanced datasets where the majority class dominates (Joshi et al., 2019). Precision and recall offer more meaningful insights, especially when minimizing false positives (e.g., wrongly predicting loyal customers as churners) or false negatives (e.g., failing to detect actual churners) is a priority. The F1-score balances these two dimensions, providing a holistic view of model effectiveness. AUC-ROC is widely favored for its ability to visualize trade-offs between true positive and false positive rates across thresholds. In business applications, economic evaluation metrics such as lift, profit curves, and cost-benefit ratios are increasingly used to align model outputs with financial outcomes. Cross-validation techniques, including k-fold and stratified sampling, are employed to ensure generalizability and reduce overfitting. Hyperparameter tuning via grid search, random search, or Bayesian optimization further enhances model calibration. Benchmarking models across diverse datasets and scenarios ensures that performance claims are robust and transferable (Xiahou & Harada, 2022). As model transparency becomes a concern, tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide interpretability into black-box models.
Machine learning-based customer retention models have been widely adopted in sectors where customer engagement directly impacts recurring revenue and long-term viability. In the telecommunications industry, churn is a critical issue due to high customer acquisition costs and ease of switching providers (Kaya et al., 2018). Predictive modeling enables telecom firms to identify dissatisfied customers early and design targeted retention campaigns. In the financial services sector, banks and insurance companies utilize churn analytics to retain premium clients and minimize the cost of lost accounts. E-commerce platforms apply ML models to monitor browsing behavior, transaction history, and product reviews to anticipate customer defection (Šemrl & Matei, 2017). Subscription-based services like streaming platforms, fitness apps, and SaaS providers rely heavily on churn prediction to improve user retention through personalization and proactive customer service (Moro et al., 2015). In retail, loyalty programs and targeted promotions are driven by insights from predictive models that estimate attrition risk (Amin et al., 2016). Airlines and hospitality sectors also leverage customer retention analytics to enhance loyalty programs and service recovery efforts. The cross-sectoral application of ML for churn analysis demonstrates its versatility and underscores its strategic importance in customer relationship management (CRM) systems. Each industry tailors its predictive models to specific customer touchpoints and behavioral indicators, requiring domain-specific data integration and feature selection.

Implementing machine learning-based churn prediction models in real-world environments entails significant technical, organizational, and ethical challenges. Data quality and integration remain critical concerns, as incomplete or inconsistent records can compromise model accuracy (Jajam et al., 2023). Organizational resistance to algorithmic decision-making, often rooted in lack of technical expertise or trust in automation, can impede adoption (Jahan & Sanam, 2024). Ensuring model scalability across customer segments, geographies, and platforms requires robust IT infrastructure and data governance frameworks. Regulatory requirements such as the General Data Protection Regulation (GDPR) in Europe impose constraints on the use of personal data, necessitating privacy-preserving techniques like data anonymization, differential privacy, and federated learning. Furthermore, algorithmic bias and fairness remain pressing issues, as models trained on skewed historical data may replicate or amplify discrimination. Transparent model deployment is increasingly expected, especially in high-stakes domains like finance or healthcare, where decisions must be auditable and explainable (Singh et al., 2023). Maintenance of ML models over time—often referred to as model drift or concept drift—requires continuous monitoring and retraining to reflect changing customer behavior and market dynamics. These considerations underscore the multidimensional nature of churn prediction model deployment, which spans technical optimization, organizational readiness, and regulatory compliance.
The primary objective of this systematic literature review is to explore and synthesize the breadth of scholarly work and practical applications surrounding the use of machine learning models for customer retention, with a specific focus on predictive analytics and churn analysis. This study aims to examine how various machine learning techniques have been employed across industries to identify patterns of customer disengagement and forecast the likelihood of churn. By mapping out the types of algorithms used—such as decision trees, support vector machines, random forests, gradient boosting, and deep learning—the review seeks to understand their relative effectiveness, limitations, and suitability for different datasets and business contexts. Another core objective is to evaluate the role of feature engineering in enhancing model performance, investigating how behavioral, transactional, demographic, and sentiment-based variables contribute to the accuracy of churn predictions. The study also aims to compare evaluation metrics used in model validation, such as precision, recall, and AUC-ROC, to assess how different approaches measure performance and align with business goals. Furthermore, this review intends to highlight common challenges faced during model deployment, including issues of data imbalance, interpretability, and regulatory compliance. By systematically reviewing peer-reviewed academic articles, case studies, and industry reports, the study endeavors to identify methodological gaps, practical limitations, and underexplored areas in the field. It also aims to provide a cross-sectoral analysis, comparing how customer retention models are tailored across domains such as telecommunications, banking, retail, and digital platforms. Overall, the objective is to offer a consolidated knowledge base that supports academic inquiry, guides industry practitioners in adopting data-driven retention strategies, and informs the development of more sophisticated, accurate, and ethically responsible churn prediction frameworks using machine learning.

LITERATURE REVIEW

Customer retention is a foundational pillar of sustainable business strategy, and the rapid growth of digital data has propelled machine learning (ML) to the forefront of retention analysis and churn prediction. The literature exploring this convergence spans diverse fields including marketing analytics, data science, artificial intelligence, and customer relationship management. This section systematically reviews existing academic contributions and practical insights on the application of ML in identifying, analyzing, and responding to customer churn. It presents a rigorous synthesis of theoretical models, algorithmic approaches, industry applications, data challenges, evaluation frameworks, and ethical considerations. The review begins by tracing the conceptual foundations of customer retention and churn analysis, laying the groundwork for understanding their strategic importance across sectors. It then examines traditional churn prediction models and their evolution into modern machine learning frameworks. Particular attention is given to algorithmic advancements—ranging from supervised learning classifiers to deep neural networks—and the role of feature engineering and dataset design in enhancing predictive accuracy. This section also evaluates performance metrics commonly used to benchmark ML models, while offering cross-sectoral comparisons of implementation practices. Finally, the literature review addresses the pressing issues of interpretability, fairness, and compliance that emerge when deploying ML in real-

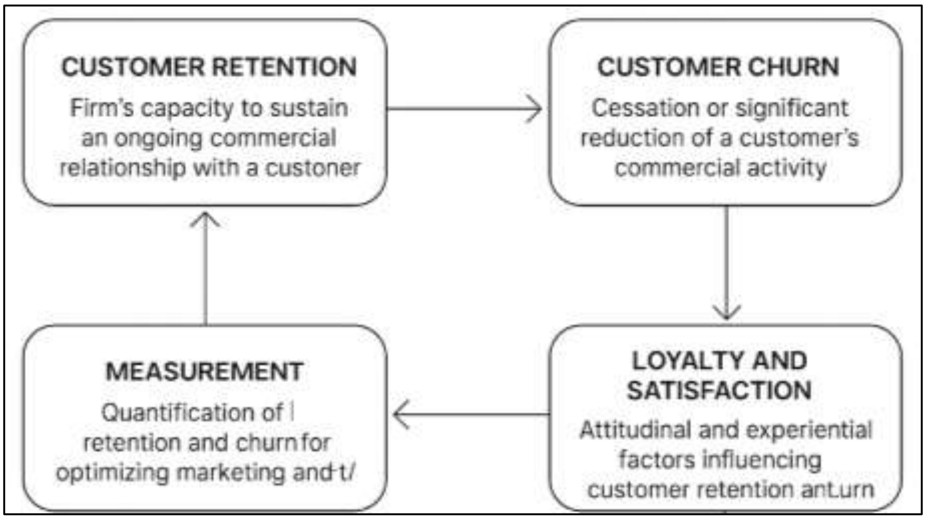
Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

world environments. Through this multi-dimensional review, the study identifies key developments, methodological gaps, and critical debates shaping the field.

Customer Retention and Churn

Customer retention is commonly defined as the firm's capacity to sustain an ongoing commercial relationship with a buyer over time, thereby maximizing the economic value derived from that customer's repeat patronage. Early work by Kassem et al. (2020) framed retention as a strategic lever that can dramatically increase profitability because retained customers generate escalating cash flows while requiring lower acquisition outlays. Subsequent conceptual refinements have emphasized a relational perspective, arguing that retention reflects both behavioral continuance (e.g., repeat purchases) and attitudinal commitment to a provider (Panimalar & Krishnakumar, 2023). Lu et al. (2014) integrated these views within their Customer Equity framework, positioning retention—alongside acquisition and add-on selling—as a core driver of lifetime value. Swetha and Dayananda (2023) further quantified retention's financial salience by linking a one-point increase in retention rate to proportional gains in firm valuation. Agrawal et al. (2018) highlighted that retention is not solely a marketing outcome but an emergent state influenced by satisfaction, perceived switching costs, and prior relationship investments. Chong et al. (2023) offered an early taxonomy of defection triggers, demonstrating that service failures, price increases, and convenience barriers erode retention in service industries. More recent meta-analyses affirm the universal relevance of retention across contexts such as banking, telecoms, and e-commerce, yet underscore the contingency of its antecedents on culture, purchase frequency, and competitive intensity. Collectively, this body of literature positions retention as a multidimensional construct encompassing transactional persistence, emotional bonds, and calculative assessments of value, thereby setting the foundation for advanced analytical efforts aimed at predicting and managing customer tenure.

Figure 4: Customer Retention, Churn, and Their Interrelationships



Customer churn—the behavioral manifestation of retention failure—is typically defined as the cessation or significant reduction of a customer's commercial activity within a predefined observation window. Scholars distinguish between voluntary churn, where customers actively switch suppliers, and involuntary churn, resulting from events such as credit default or service termination (Awasthi, 2022). Amin et al. (2023) further categorize churn into complete, partial, and hidden forms, noting that customers may maintain nominal accounts while diverting the majority of their spending elsewhere. (Sari et al., 2023) introduced the stochastic "buy-till-you-die" framework, treating churn as an unobserved lifetime horizon that can be inferred from repeat-purchase patterns. Mouli et al., (2024) emphasized that the operational definition of churn must align with managerial decision cycles; for instance, a telecommunications firm may label a prepaid subscriber inactive after 90 days of zero top-ups, whereas a bank might allow a full fiscal year before deeming an account dormant. Adhikary and Gupta (2020) argued for viewing churn as a dynamic event, advocating hazard-rate models that capture time-varying covariates such as promotional exposure. Empirical studies confirm that churn drivers range from dissatisfaction with core service quality to external incentives like

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

competitor discounts or network effects. Importantly, the churn construct is sensitive to industry norms: in subscription businesses churn is binary, whereas in retail it often manifests as gradual share-of-wallet erosion. This definitional complexity necessitates precise operationalization before predictive modeling can yield actionable insights, reinforcing the need for clarity about observational windows, measurement granularity, and domain-specific churn archetypes.

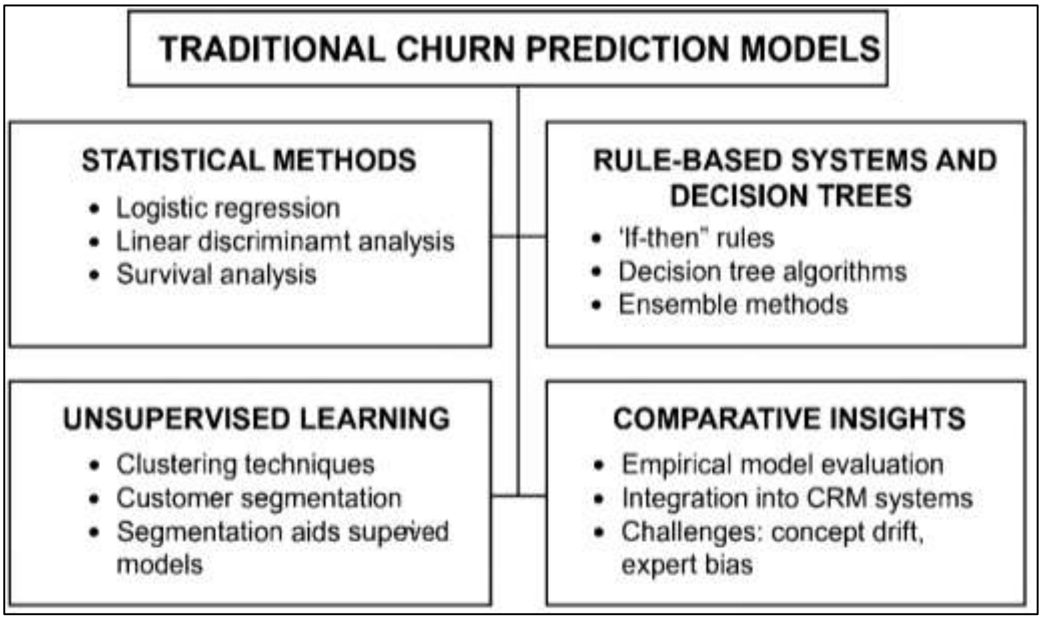
Quantifying retention and churn is imperative for allocating marketing resources and optimizing customer lifetime value (CLV). Jajam et al. (2023) offered the seminal CLV formula linking margin, retention rate, and discount factor, illustrating how even marginal improvements in retention can yield disproportionate profit gains. Sari et al. (2023) translated these principles to firm valuation, showing that Wall Street rewards companies with demonstrably superior retention profiles. Jajam et al. (2023) recommended transitioning from aggregate retention metrics to granular cohort analyses that isolate heterogeneity in tenure and spend trajectories. Contemporary analytics adopt survival analysis and hidden Markov models to estimate individual-level hazard rates and forward-looking retention probabilities, thereby enabling micro-segmented marketing interventions. From a managerial accounting perspective, retention metrics such as customer attrition rate, renewal rate, and average relationship length feed directly into forecasting dashboards used for capacity planning and revenue recognition. In regulated industries like finance, precise churn measurement also supports compliance by evidencing fair-treatment standards and mitigating conduct risk. Additionally, balanced-scorecard frameworks increasingly incorporate retention as a critical success factor alongside acquisition, cross-sell, and service quality indicators. The economic logic culminates in the recognition that retention is both a lagging indicator of past performance and a leading indicator of future cash flows, thus commanding strategic focus in resource-constrained environments. Accurate, context-sensitive measurement systems enable firms to prioritize high-risk customers for retention campaigns, evaluate program ROI, and refine predictive algorithms through feedback loops, solidifying the centrality of retention and churn constructs in data-driven customer management.

Traditional Churn Prediction Models

Before the widespread adoption of machine learning, traditional statistical models dominated the field of churn prediction, leveraging well-established methods such as logistic regression, linear discriminant analysis, and survival analysis. Logistic regression was widely adopted due to its interpretability, efficiency, and robustness in binary classification tasks such as predicting churn versus retention (Faritha Banu et al., 2022). It allowed researchers to estimate the probability of churn based on explanatory variables such as customer demographics, tenure, service usage, and billing frequency. Umayaparvathi and Iyakutti (2012) demonstrated the effectiveness of logistic regression in retail loyalty analysis, showing its utility in modeling both transactional and demographic variables. Similarly, linear discriminant analysis was used to classify customers based on their likelihood to defect, particularly when assumptions of multivariate normality and equal covariance matrices were met. Despite their limitations in modeling nonlinear relationships, these methods provided valuable baselines for predictive performance and benchmarking. Survival analysis, particularly Cox proportional hazards models, introduced a temporal dimension to churn prediction by modeling the hazard rate over time and identifying time-dependent churn risks. These approaches were particularly useful in contexts such as subscription services and telecom, where time until churn held strategic importance. While statistically grounded and computationally efficient, traditional methods were constrained by their reliance on linear assumptions, inability to capture complex interactions, and sensitivity to multicollinearity and data sparsity. Nonetheless, their transparency and ease of interpretation made them favorable among decision-makers, particularly in regulated industries such as finance and insurance (Jajam et al., 2023). These models laid the groundwork for more advanced, flexible approaches and continue to serve as benchmarks in modern churn prediction studies. Beyond statistical regressions, early churn prediction models also relied on rule-based systems and decision trees, which offered intuitive, logic-driven frameworks for classification tasks. Rule-based systems were particularly prominent in customer relationship management software, where expert-defined heuristics were encoded into "if-then" rules to flag at-risk customers (Banu et al., 2022). These systems were straightforward to implement and offered immediate interpretability, especially when customer behaviors conformed to predictable patterns.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Figure 5:Overall T raditional Churn Prediction Models



However, their rigidity made them susceptible to performance degradation in dynamic environments where customer preferences evolved rapidly. Decision trees, including algorithms such as ID3, C4.5, and CART, introduced a more data-driven and flexible approach to rule induction (Xiahou & Harada, 2022). These methods partitioned the dataset into subsets based on attribute values, forming a tree-like structure that guided classification decisions (Zhu et al., 2023). One of the strengths of decision trees lay in their ability to handle both categorical and numerical data while revealing feature importance through their structure (Cheng et al., 2019). Coussement and Kaya et al. (2018) found that decision trees performed competitively in the banking sector by capturing interaction effects among churn drivers such as loan defaults, communication patterns, and customer complaints. However, early decision trees were often prone to overfitting and required pruning to maintain generalizability. Despite these challenges, their transparency and visualization benefits made them attractive for managerial interpretation and customer profiling (Hadden et al., 2005). Ensemble methods like bagging and boosting were later developed to overcome the instability and variance associated with single-tree classifiers. In early practice, decision trees and rule-based models played a critical role in operationalizing customer segmentation and informing early warning systems in retention departments (Mouli et al., 2024). Their structured logic also made them a preferred choice for integration into enterprise dashboards and legacy CRM systems, where ease of explanation was as valuable as accuracy.

Unsupervised learning approaches such as clustering and segmentation have long been used to complement churn prediction by identifying hidden structures within customer data. Clustering algorithms, particularly k-means and hierarchical clustering, were employed to group customers based on similarities in behavior, demographics, and transactional patterns. These techniques allowed businesses to profile customers into homogenous segments such as "high risk," "stable," or "new joiners," aiding targeted retention strategies. Jajam et al. (2023) showed that cluster-based segmentation enhanced the performance of subsequent classification models by reducing intra-class variance. Kohonen Self-Organizing Maps (SOMs) were also used to visualize customer groups on a two-dimensional grid, highlighting behavioral patterns and churn-prone segments. In practice, unsupervised approaches were particularly useful in the early stages of churn analytics when labeled churn data were scarce or inconsistent. Although these methods did not directly predict churn probabilities, they informed the design of supervised models by revealing variable groupings and facilitating feature selection. For example, Latent Class Analysis (LCA) helped identify customer archetypes with differential attrition risks, allowing firms to design differentiated loyalty programs. Nonetheless, clustering models faced challenges related to determining optimal cluster numbers and ensuring segment stability over time. Another limitation was their reliance on distance metrics, which were sensitive to scaling and required substantial preprocessing. Despite these constraints,

Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

segmentation remained an indispensable tool in traditional churn modeling frameworks, particularly when used in conjunction with decision trees and logistic regressions for a more holistic customer insight.

The effectiveness of traditional churn prediction models has been a focus of empirical comparison in both academic and applied research. Banu et al.(2022) compared logistic regression, neural networks, and decision trees in predicting churn in the cellular services sector, finding modest differences in performance but significant variations in interpretability and operational utility. Janssens et al. (2022) conducted a large-scale empirical comparison of predictive models across six organizations and found that traditional statistical models were competitive with more complex algorithms in environments where customer behavior was stable and well-structured. However, in dynamic markets or datasets with nonlinear patterns, traditional models underperformed relative to newer methods. Zhu et al.(2023) highlighted that the performance gap between logistic regression and advanced models often narrowed when proper feature engineering and class imbalance handling were applied. In terms of practical deployment, traditional models were widely integrated into early customer relationship management (CRM) systems, allowing firms to make tactical decisions on loyalty campaigns, retention budgets, and personalized offers. Their computational efficiency enabled real-time scoring on large datasets, making them viable in telecom call centers and retail systems where immediate churn signals were needed. However, these models lacked adaptability to concept drift—changes in customer behavior over time—requiring frequent recalibration and manual parameter adjustments. Additionally, reliance on expert-selected features sometimes introduced bias and omitted latent behavioral variables. Still, their simplicity and ease of interpretation made them enduring tools in many organizations, especially where explainability and regulatory transparency were critical. As a result, traditional churn models served as foundational approaches that informed the structure, evaluation, and expectations for more complex machine learning frameworks that followed.

Emergence of Machine Learning in Retention Analytics

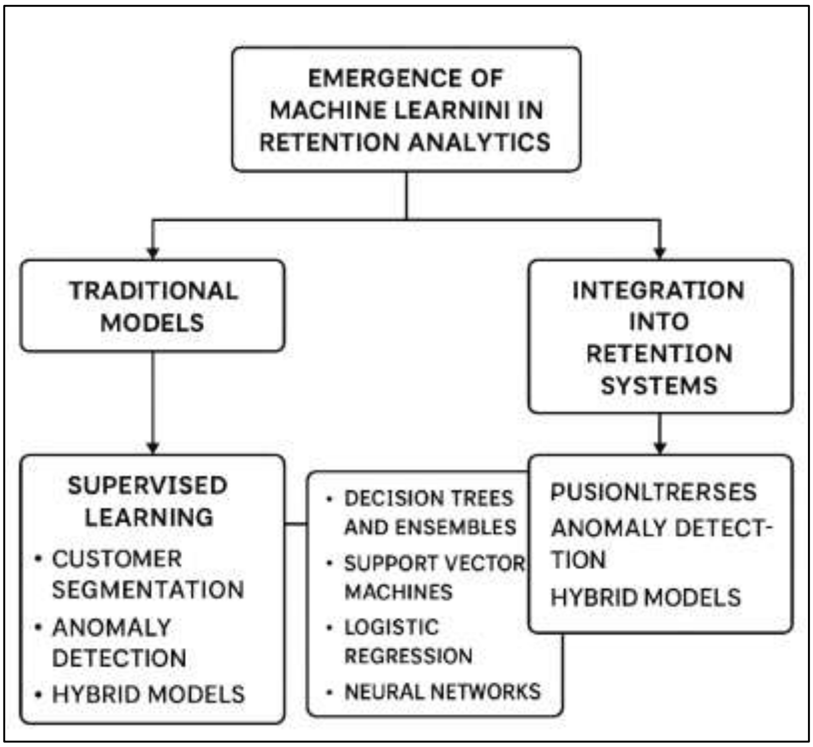
The growing complexity and volume of customer data have necessitated a paradigm shift from traditional statistical methods to more flexible and adaptive machine learning (ML) techniques in customer retention analytics. Traditional models such as logistic regression and discriminant analysis, while useful for linear and interpretable relationships, lack the capability to model nonlinear interactions or discover complex, high-dimensional patterns. As datasets became richer with behavioral, transactional, and unstructured text data, ML algorithms began to outperform classical techniques in both predictive accuracy and feature discovery (Cheng et al., 2019). The rise of customer-centric data from digital platforms catalyzed this shift, where retention strategies demanded real-time and personalized insights across multiple touchpoints (Mouli et al., 2024). Machine learning enables automated learning from evolving datasets, making it particularly useful in dynamic customer environments such as telecom, banking, and e-commerce. Algorithms such as support vector machines (SVM), random forests, and gradient boosting machines have shown higher adaptability and robustness in modeling churn. Unlike rule-based models that rely on predefined thresholds, ML algorithms continuously adjust to new patterns, accommodating concept drift and complex customer behaviors. Moreover, advancements in computing power and data infrastructure, including cloud-based analytics and GPU-accelerated training, have facilitated the large-scale deployment of ML-based retention models in both operational and strategic decision-making contexts. This transition represents a significant evolution in the analytical capabilities available to firms aiming to reduce churn and foster long-term customer engagement.

Supervised learning remains the dominant approach in machine learning–based churn prediction, where historical labeled data—consisting of churned and retained customers—are used to train models that generalize to unseen instances. Decision trees and their ensemble variants, such as random forests and gradient boosting machines (GBMs), have consistently demonstrated high predictive performance in identifying churn-prone customers across domains like telecom, retail, and finance (Jajam et al., 2023; Zhu et al., 2023). Decision trees split data recursively based on feature importance, creating interpretable decision paths, while random forests aggregate multiple trees to reduce variance and improve generalizability (Sari et al., 2023). GBMs further improve predictive power by sequentially correcting errors from previous iterations, making them ideal for capturing subtle churn indicators. Support vector machines (SVMs) have been applied for high-dimensional datasets, using kernel functions to classify nonlinear patterns in customer behavior. Logistic regression

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

with regularization (LASSO or Ridge) is still utilized as a benchmark in ML studies, particularly when interpretability is needed. Neural networks, particularly multilayer perceptrons (MLPs), have shown strong performance when data complexity requires deep, hierarchical learning structures. These supervised algorithms are trained using a range of evaluation metrics—such as accuracy, recall, and AUC-ROC—to ensure robustness across class imbalance challenges. Studies have demonstrated that ML algorithms consistently outperform traditional statistical methods in predictive accuracy, especially when engineered features and cross-validation techniques are effectively applied. The scalability and flexibility of these supervised techniques have enabled organizations to implement churn models in real-time decision systems, driving personalized retention strategies at scale.

Figure 6: Framework Illustrating the Emergence of Machine Learning in Customer Retention Analytics



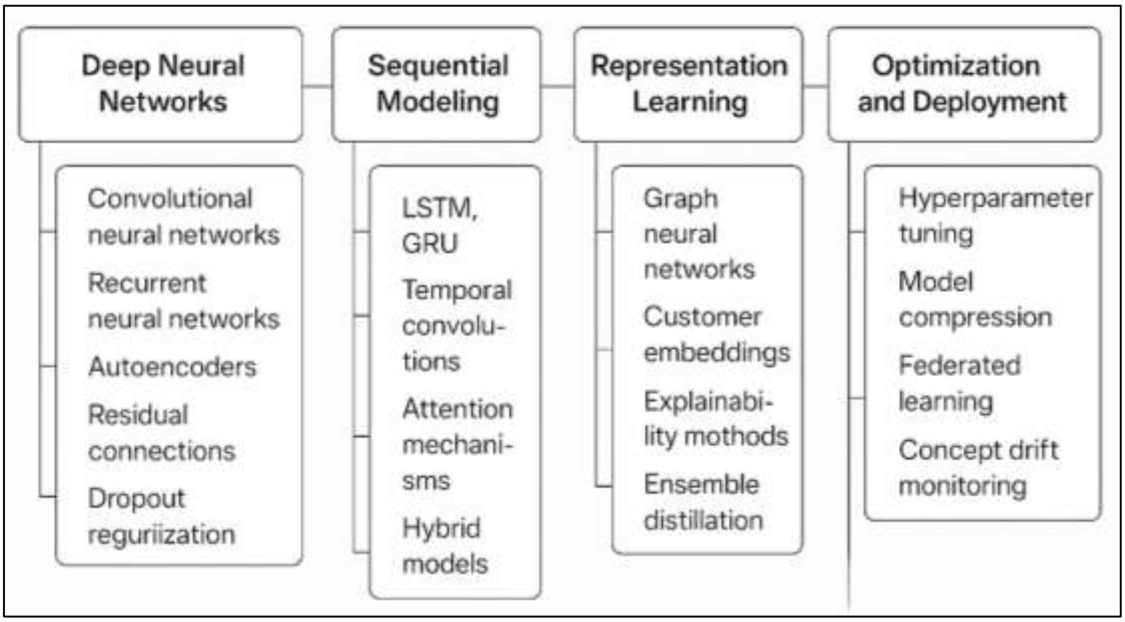
The practical integration of machine learning into enterprise-level retention systems has been facilitated by advancements in model automation, explainability, and deployment frameworks. Companies now utilize end-to-end ML pipelines that automate data preprocessing, feature selection, model training, and prediction scoring within CRM systems (Kaya et al., 2018). AutoML tools streamline model selection and hyperparameter tuning, enabling non-expert users to deploy highly accurate models with minimal manual intervention. These tools often incorporate ensemble techniques and performance validation protocols, reducing the risk of overfitting and poor generalization. Cloud-based platforms such as AWS SageMaker, Google Cloud AI, and Microsoft Azure ML have further lowered the technical barriers to deploying retention models at scale, providing robust environments for model training, inference, and monitoring. Explainability tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have addressed the "black-box" concerns traditionally associated with advanced ML models, offering transparency into decision-making processes and satisfying regulatory compliance standards. Model deployment in churn analytics has been successful across various sectors. In telecommunications, predictive ML models are embedded into customer service systems to prompt real-time retention offers. E-commerce platforms use ML-based personalization engines to dynamically recommend products and re-engage high-risk users. In banking, real-time credit risk and churn scoring models are integrated with mobile apps and call centers to improve customer service outcomes (Brito et al., 2024). The convergence of ML, cloud infrastructure, and business intelligence systems has thus operationalized retention analytics, enabling continuous learning, campaign optimization, and seamless scaling of predictive insights within organizational workflows.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Deep Learning and Advanced Techniques in Churn Prediction

The application of deep learning to churn prediction began as firms amassed high-granularity clickstream, usage, and sentiment data that exceeded the representational capacity of shallow models. Multilayer perceptrons quickly gave way to deeper convolutional and recurrent architectures capable of extracting hierarchical abstractions from raw behavioral logs (Saha et al., 2024). Convolutional neural networks (CNNs), originally popularized for image recognition, were repurposed to capture local activation patterns in sequential service-usage matrices, enabling telcos to detect subtle frequency shifts preceding defection. Autoencoders further advanced representation learning by compressing high-dimensional transaction vectors into dense latent codes that preserved purchase periodicity and monetary value while filtering noise. Studies comparing CNN-based churn detectors with gradient boosting machines reported up to ten-point gains in AUC-ROC when unstructured log data were available, underscoring the advantage of end-to-end feature extraction. Dropout regularization and batch normalization mitigated overfitting and accelerated convergence, making deep models viable on enterprise hardware. Moreover, residual connections popularized by ResNet variants allowed deeper stacks without vanishing gradients, facilitating richer temporal abstractions in month-over-month customer journeys (Agrawal et al., 2018). Collectively, these advances enabled practitioners to ingest raw event streams, loyalty scores, and textual complaints into a unified neural pipeline, yielding granular churn probabilities at scale. The literature consistently demonstrates that when data volume and variety are high, deep neural architectures surpass traditional and shallow learners by discovering nonlinear cues—such as session volatility or complaint tone shifts—that precede observable disengagement.

Figure 7: Comprehensive Framework of Deep Learning and Advanced Techniques



Because churn is inherently a temporal phenomenon, recurrent neural networks (RNNs) and their gated extensions have gained prominence for modeling sequential customer behavior. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, designed to capture long-range dependencies without vanishing gradients, have been successfully applied to telecom call-detail records and subscription media logs, achieving superior recall of early-stage churners compared with fixed-window classifiers (Khattak et al., 2023). Temporal Convolutional Networks (TCNs) further enhanced sequential modeling by replacing recurrence with dilated convolutions, yielding parallelization benefits and deeper receptive fields (Shoja & Tabrizi, 2019). Attention mechanisms—initially proposed for neural machine translation—have been adapted to weigh salient events such as service outages or billing anomalies, thereby improving interpretability and boosting F1-scores in churn detection (Saha et al., 2023). Hybrid models that fuse LSTM backbones with attention layers demonstrate notable gains in precision when predicting voluntary versus involuntary attrition, as the attention weights highlight contextual triggers like competitor promotions captured in social feeds.

Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Hierarchical recurrent networks further disaggregate sequences into session-level and customer-lifecycle tiers, capturing micro-bursts of dissatisfaction nested within macro usage trends. Comparative studies across banking and e-commerce datasets reveal that sequence-aware architectures can reduce false-negative churn predictions by up to 30 percent relative to static feature models. These findings underscore the value of modeling temporal context explicitly, allowing organizations to intervene during critical windows when retention actions are most effective.

Advanced churn frameworks increasingly rely on sophisticated representation learning techniques that extend beyond vanilla feed-forward designs. Variational autoencoders have been leveraged to generate smooth latent manifolds of customer behavior, facilitating anomaly detection for early churn warnings (Saha et al., 2023). Embedding layers derived from word2vec and doc2vec models convert categorical attributes—such as tariff plans or product categories—into dense vectors that encode relational similarity, thereby enhancing downstream classification performance (Wanganga & Qu, 2020). Graph Neural Networks (GNNs) model customer interactions and social influence effects by treating users as nodes and communication ties as edges; empirical evidence shows that incorporating network centrality measures improves churn AUC by capturing peer contagion (Utku & Akcayol, 2020). Capsule networks, though nascent, have been explored to preserve hierarchical part–whole relationships in multichannel behavioral data, yielding promising early results in subscription media churn tasks (Jajam et al., 2023). Alongside predictive power, explainability tools such as SHAP (Wanganga & Qu, 2020) and LIME have become integral to deep churn pipelines, providing customer-level rationales that satisfy regulatory scrutiny and support agent scripting. Layer-wise relevance propagation further visualizes contribution scores across time steps and channels, helping managers trace churn signals to distinct events like declined payments or negative reviews (Agrawal et al., 2018). Ensemble distillation techniques compress complex architectures into interpretable surrogates without significant accuracy loss, bridging the gap between advanced modeling and managerial usability. These developments collectively demonstrate that cutting-edge representation learning and interpretability frameworks can coexist, yielding high-fidelity and transparent churn predictions.

Feature Engineering and Data Preparation Strategies

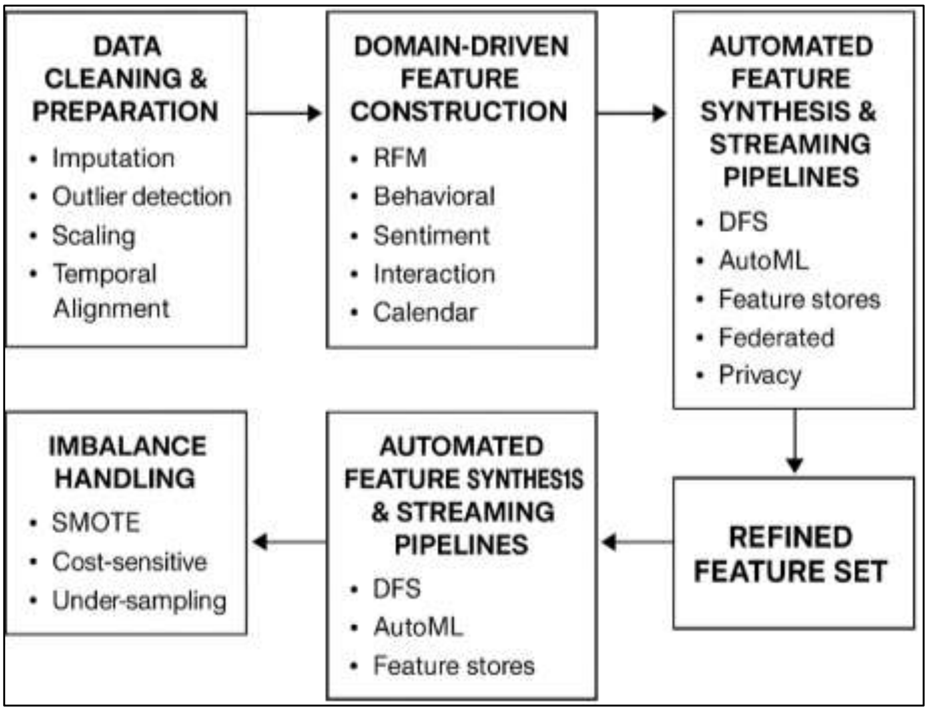
Feature engineering for churn prediction begins with rigorous data preparation, as raw customer data often contain noise, missing values, and temporal inconsistencies that degrade model accuracy. Early studies stressed the impact of data quality on predictive performance, showing that imputed datasets improved lift scores by up to 15 percent compared with case-wise deletion (Shoja & Tabrizi, 2019). Common imputation techniques range from mean substitution and k-nearest neighbor algorithms to multiple imputation by chained equations, each selected according to the mechanism of missingness diagnosed via Little's MCAR test (Saha et al., 2023). Outlier detection using Mahalanobis distance has been shown to reduce false-positive churn alerts in banking datasets, while robust z-score scaling stabilizes gradient descent in neural churn models trained on highly skewed monetary features. Temporal alignment is equally critical; Shaaban et al. (2012) demonstrated that synchronizing billing, usage, and support-ticket logs to a common weekly granularity increased AUC-ROC by six points in telecommunications data. Seasonality adjustments via STL decomposition mitigate periodic fluctuations that otherwise obscure underlying defection trends. Transactional sequences are often converted into lagged variables—such as 7-, 30-, and 90-day usage averages—capturing short- and long-term behavioral signals (Wanganga & Qu, 2020). Moreover, logarithmic and Box-Cox transformations help normalize heavy-tailed purchase distributions, facilitating distance-based learning algorithms (Garimella et al., 2021). Collectively, these preprocessing techniques form the foundation for reliable feature extraction by ensuring that subsequent modeling phases are insulated from artifacts arising from data sparsity, heterogeneity, and scale disparities.

Domain expertise plays a decisive role in crafting high-signal predictors for churn analytics. The classic Recency, Frequency, Monetary (RFM) framework has been adapted across sectors, where recency may represent days since last login, frequency denotes session counts, and monetary captures cumulative spend (Wanganga & Qu, 2020). Extensions incorporate engagement intensity—such as average minutes per session—and diversity indices measuring breadth of product usage, which (Saha et al., 2024) found to improve telecom churn recall by 12 percent. Incorporating billing attributes like payment method volatility and overdue balance flags has proven effective in credit-

Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

card attrition studies. Behavioral micro-events, including click-stream dwell time and scroll depth, have been embedded as aggregate statistics or time-series signatures using Fast Fourier Transform coefficients to capture cyclical engagement. Sentiment features extracted via lexicon-based or transformer-based language models from support tickets and social-media conversations add qualitative context to quantitative usage variables. Cross-feature interactions—such as tenure × complaint frequency—are engineered to highlight conditional churn risks, echoing findings from nested logit models in service failure research. Calendar features capturing renewal anniversaries, public holidays, or end-of-quarter sales cycles also contribute explanatory power, particularly in subscription businesses. When orchestrated thoughtfully, domain-driven features transform raw records into semantically meaningful constructs that align with managerial levers, enabling precise intervention design and cost-efficient retention campaigns.

Figure 8: Overall Feature Engineering and Data Preparation Strategies



Once candidate variables are constructed, feature selection mitigates redundancy, multicollinearity, and overfitting. Filter methods—such as mutual information and chi-square tests—provide fast, model-agnostic rankings that have proven effective in pruning thousands of web-log attributes down to a manageable subset without sacrificing accuracy. Wrapper techniques like recursive feature elimination employ classifier feedback loops; Agrawal et al. (2018) reported that coupling RFE with support vector machines yielded a 5-point F1-score improvement on e-commerce churn. Embedded approaches integrated within tree-based ensembles exploit intrinsic split criteria to surface high-gain splits, as demonstrated by Utku and Akcayol (2020) and later extended in gradient boosting frameworks (Jajam et al., 2023). Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding compress high-dimensional behavioral spectra into orthogonal components, aiding downstream K-means clustering for risk segmentation (Utku & Akcayol, 2020). Meanwhile, autoencoder-based bottlenecks learn nonlinear embeddings tailored to deep classifiers, outperforming linear reductions on heteroskedastic spending data (Zhang et al., 2023). Addressing severe class imbalance—common when churners constitute < 10 percent of the base—remains essential; Synthetic Minority Over-sampling Technique (SMOTE) and its adaptive variants boost minority representation, while cost-sensitive learning penalizes false negatives more heavily. Ensemble under-sampling, exemplified by EasyEnsemble, combines balanced bootstrap samples to maintain decision boundary diversity. These strategies collectively refine the feature space, promoting stable generalization across temporal, geographic, and product cohorts.

Recent advances in automation have reduced the manual burden of feature engineering while expanding feature search spaces. Deep Feature Synthesis (DFS) algorithms automatically generate higher-order aggregates—such as rolling-window variances or ratio metrics—by traversing entity-relationship graphs, delivering lift gains of up to eight percentage points in SaaS churn competitions. AutoML platforms extend this automation by coupling DFS with algorithm selection and hyperparameter tuning, democratizing advanced analytics for non-expert practitioners. In real-time retention systems, streaming frameworks such as Apache Kafka and Spark Structured Streaming maintain incremental feature stores that update RFM counters and sentiment scores within seconds, enabling proactive churn interventions (Panimalar & Krishnakumar, 2023). Feature stores standardize definitions, versioning, and lineage, facilitating reproducible experiments and compliant audits. Federated feature engineering pushes computations to edge devices, allowing mobile usage vectors to be aggregated locally and encrypted gradients returned centrally, thereby meeting GDPR's data-minimization principle. Differential privacy noise injected into feature aggregations further shields individual-level attributes while retaining cohort-level signal. Explainability overlays—using SHAP interaction values—quantify how engineered features drive individual churn scores, supporting transparent customer-service conversations and regulator inquiries. Together, automated synthesis, low-latency pipelines, and privacy-aware mechanisms represent a holistic evolution in feature engineering, ensuring that sophisticated predictive insights coexist with operational agility and ethical data stewardship.

Metrics for Churn Prediction Models

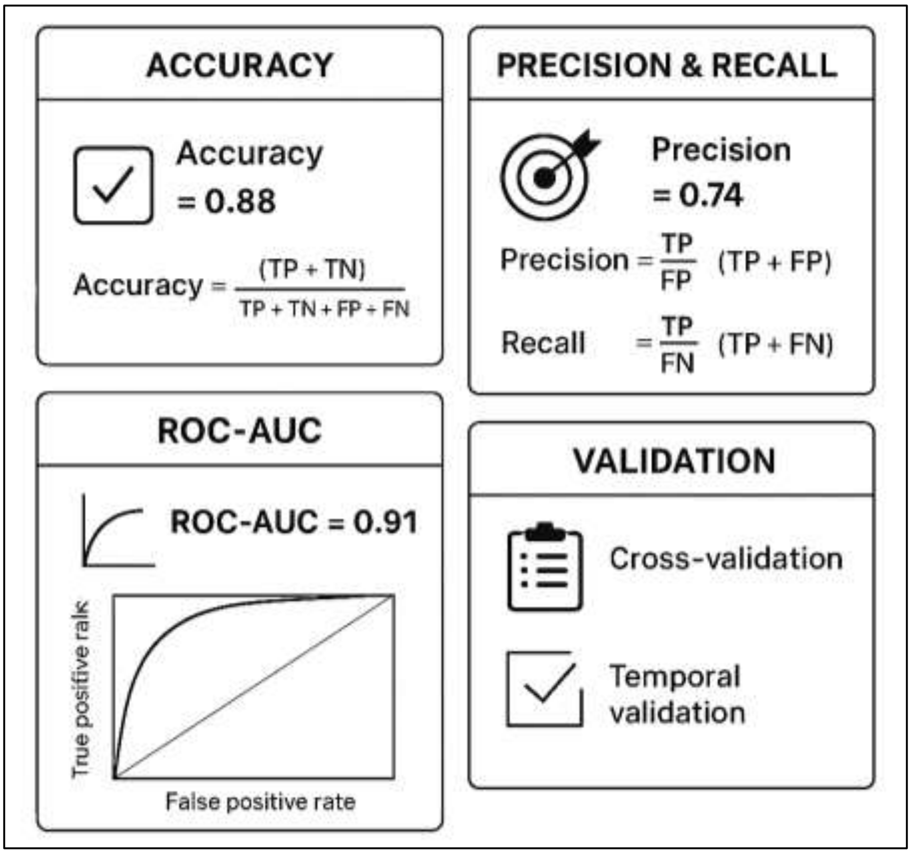
Churn prediction models are typically evaluated using standard classification metrics, with accuracy being one of the most commonly reported. However, accuracy alone can be misleading, particularly in imbalanced datasets where non-churners vastly outnumber churners (Amin, Al-Obeidat, et al., 2019; Subrato, 2018). In such cases, a model that predicts all customers as non-churners may still yield high accuracy, despite providing no practical value. Precision and recall offer more nuanced insights into model performance. Precision measures the proportion of predicted churners who actually churned, making it crucial when the cost of a false positive is high—such as offering discounts to customers unlikely to leave (Abdullah Al et al., 2022; Mishra & Reddy, 2017). Recall, on the other hand, measures the proportion of actual churners correctly identified, which is vital in settings where missing a potential churner is costlier than issuing unnecessary retention offers (Amin, Shah, et al., 2019; Ara et al., 2022). The F1-score harmonizes precision and recall into a single metric, balancing the trade-offs and offering a clearer picture when both false positives and false negatives carry business risks. Haddadi et al. (2024) emphasized that F1-score outperforms accuracy in evaluating models on imbalanced customer datasets. Abdulsalam et al. (2022) noted that decision tree and ensemble models exhibit variable performance across these metrics, necessitating a multi-metric approach for robust evaluation. Furthermore, Coussement and Bock (2013) recommended evaluating these metrics under different churn thresholds, as class boundaries often vary across industries. Such granularity ensures that model outputs align with stakeholder priorities and marketing resource constraints (Rahaman, 2022).

Receiver Operating Characteristic (ROC) curves and their corresponding Area Under the Curve (AUC) scores are widely used to assess classifier performance across different probability thresholds. The ROC curve plots the true positive rate (recall) against the false positive rate, providing a comprehensive visualization of model performance independent of class distribution (Abdulsalam et al., 2022; Masud, 2022). AUC-ROC values closer to 1.0 indicate superior discrimination capability, while values near 0.5 suggest random guessing. AUC has been frequently used in benchmarking studies to compare models such as logistic regression, random forests, and gradient boosting machines across datasets (Khodabandehlou & Rahman, 2017; Sazzad & Islam, 2022). However, ROC analysis can become less informative in highly imbalanced datasets, where Precision-Recall (PR) curves are preferred (Saito & Rehmsmeier, 2015). Cost-sensitive evaluation adds another critical layer by incorporating the economic implications of prediction errors. (Amin, Shah, et al., 2019) proposed profit-based metrics where the cost of false positives (unnecessary retention incentives) and false negatives (missed churners) are modeled explicitly. Wagh et al. (2024) developed a cost-benefit matrix to calculate expected retention value, considering campaign costs and recovered revenues. Huang et al. (2012) suggested incorporating business constraints directly into model optimization through cost-weighted loss functions. Such approaches align model evaluation with actionable marketing outcomes. Moreover, hybrid metrics such as lift and gain charts compare model

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

performance relative to random targeting and help assess the concentration of true churners in the top-ranked predictions (Panimalar & Krishnakumar, 2023; Shaiful et al., 2022). These methods are essential for campaign planning, where firms may only target a fixed percentage of the customer base. Collectively, ROC-AUC, PR curves, and cost-sensitive metrics provide a more practical and economically grounded evaluation framework for churn modeling (Adar & Md, 2023; Akter & Razzak, 2022).

Figure 9: Key Metrics and Validation Techniques for Evaluating Churn Prediction Models



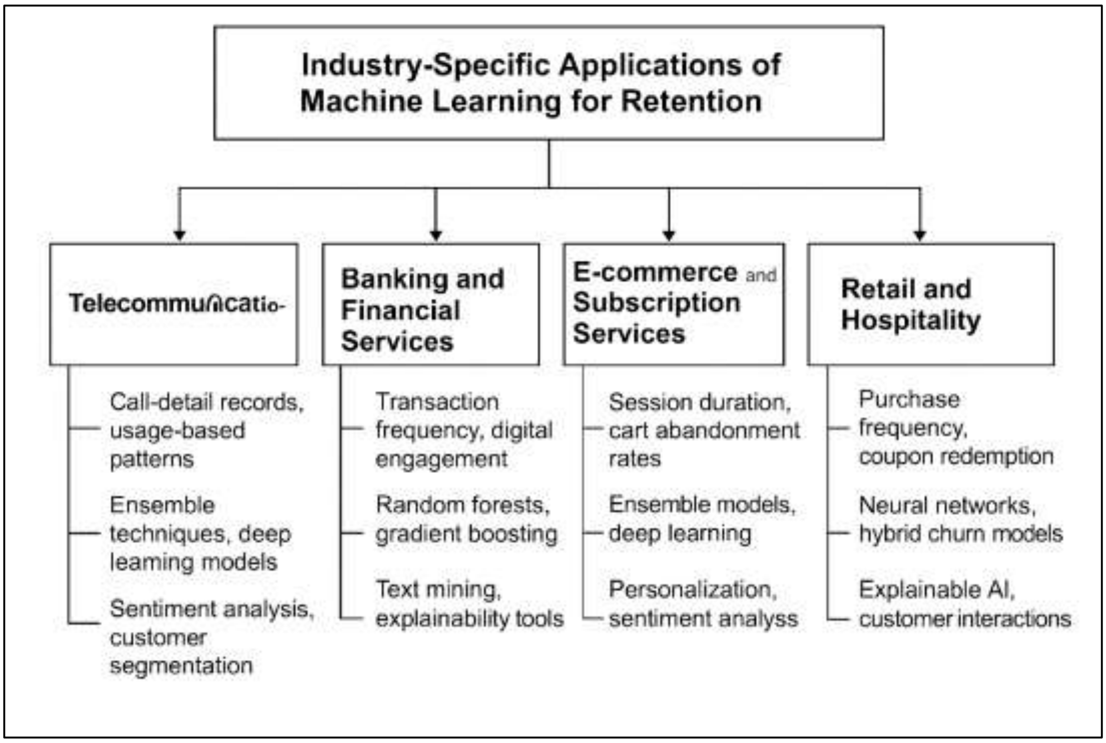
Beyond discrimination and classification accuracy, calibration measures the alignment between predicted probabilities and observed outcomes, which is essential for decision-making based on risk scores (Qibria & Hossen, 2023; Maniruzzaman et al., 2023; Akter, 2023). Well-calibrated models assign probabilities that reflect true churn likelihoods—for instance, customers with a predicted churn risk of 0.70 should, on average, churn 70% of the time. Reliability diagrams and Brier scores are commonly used to assess calibration (Masud, Mohammad, & Ara, 2023; Masud, Mohammad, & Sazzad, 2023); lower Brier scores indicate higher probability accuracy. Poor calibration can result in misallocation of retention resources, where low-risk customers are targeted and high-risk customers are overlooked. Haddadi et al. (2024) emphasized that ensemble models often require post-hoc calibration using methods like Platt scaling or isotonic regression to ensure probabilistic validity. Lift charts, another vital metric in churn prediction, measure the ratio of positive outcomes captured by the model to those expected by random selection. For instance, a lift of 3 in the top decile indicates that targeting the top 10% of customers by predicted churn probability yields three times more churners than random targeting (Vijaya & Sivasankar, 2018). Gain charts, which cumulatively plot true churns captured across percentile groups, provide a visual understanding of how quickly models capture churners. Decile-based or quantile reporting further supports practical deployment by segmenting customers into score-based buckets, enabling marketers to define tiered interventions (Haddadi et al., 2024; Shamima et al., 2023). These evaluation tools are critical for prioritizing high-risk segments in constrained campaign settings and for communicating model value to non-technical stakeholders. Together, calibration, lift, and decile analysis enhance the interpretability and actionability of churn prediction models, especially when integrated into customer relationship management systems. Evaluating churn prediction models also requires rigorous validation to ensure performance generalizability and stability over time(Ashraf & Ara, 2023). Cross-validation, particularly k-fold and stratified k-fold techniques, is extensively used to partition data into training and test sets while preserving class balance. This ensures that performance metrics are not inflated due to overfitting or chance splits. Nested cross-validation is often adopted when both model selection and hyperparameter tuning are involved, thereby mitigating selection bias. Temporal validation techniques are especially important for churn analytics, where customer behavior evolves over time. Time-based splits—where models are trained on past data and validated on future data—provide a more realistic estimation of out-of-sample performance (Ahmad et al., 2019; Sanjai et al., 2023). Rolling-window validation, used in subscription-based services, mimics real deployment scenarios where models are retrained periodically using the most recent data. Stability across segments, such as demographics or regions, is also assessed using subgroup AUC or fairness metrics to ensure consistent performance (Amin, Al-Obeidat, et al., 2019; Tonmoy & Arifur, 2023). Sensitivity analyses evaluate how metric outcomes shift under different feature sets or modeling assumptions, offering insights into model robustness. Moreover, confidence intervals and bootstrapping are used to quantify uncertainty in performance estimates, especially in small sample environments (Razzak et al., 2024; Coussement & Bock, 2013). These validation techniques are essential not only for technical model selection but also for operational deployment, where model failure can lead to reputational and financial losses. Comprehensive validation ensures that churn prediction models remain reliable, equitable, and actionable across deployment cycles and customer segments (Hossain, Haque, et al., 2024; Hossain, Yasmin, et al., 2024).

Industry-Specific Applications of Machine Learning for Retention

The telecommunications industry has been at the forefront of applying machine learning for customer retention due to high customer acquisition costs and intense market competition. Churn is particularly problematic in this sector as switching barriers are low and customers are often influenced by price promotions and network performance. Early work by Wagh et al. (2024) applied neural networks to call-detail records, achieving better churn detection than traditional regression models. Vijaya and Sivasankar (2018) employed decision trees, support vector machines (SVM), and logistic regression, finding that tree-based models provided interpretable and accurate churn predictions. More recent applications have incorporated ensemble techniques such as gradient boosting and random forests, improving the AUC-ROC by capturing nonlinear interactions among features like average call duration, SMS volume, and billing inconsistencies (Akter & Shaiful, 2024; Subrato & Md, 2024; Wagh et al., 2024). Deep learning models, particularly convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, have been used to process temporal telecom data and identify usage-based churn patterns. These models allow for real-time churn scoring and the dynamic deployment of retention offers through CRM platforms. Additionally, sentiment analysis of customer complaints and social media content has been integrated to detect dissatisfaction signals before they escalate to defection (Amin, Al-Obeidat, et al., 2019; Akter, 2025; Md et al., 2025). Studies have also shown the effectiveness of customer segmentation through unsupervised learning to tailor retention strategies by user group. With streaming data pipelines and real-time inference, telecom companies can proactively intervene within moments of detecting churn risk, turning predictive analytics into operational retention levers (Islam & Debashish, 2025; Islam & Ishtiaque, 2025; Mirkovic et al., 2022).

The banking and financial services industry has increasingly adopted machine learning models to combat attrition, particularly among high-net-worth and digitally engaged customers. Churn in this sector often manifests through dormancy, balance erosion, or migration to competing institutions. (Haddadi et al., 2024) employed survival analysis and logistic regression to predict tenure-based defection, while more recent studies adopted random forests and gradient boosting machines for modeling multifactorial churn determinants such as transaction frequency, loan defaults, and credit card usage patterns (Ahmad et al., 2019; Akter, 2025). Autoencoders and deep neural networks have been utilized to learn customer representations from high-dimensional behavioral logs, such as ATM usage, digital wallet interactions, and mobile app navigation sequences (Amin et al., 2016). These approaches outperform linear classifiers by identifying nonlinear relationships between features like account tenure and digital engagement. Recurrent neural networks have been applied to sequential transaction histories, allowing models to anticipate churn during financial inactivity or periods of declining creditworthiness. Text mining techniques have also been integrated into models through natural language processing (NLP) applied to customer service transcripts and feedback forms, enriching the feature space with sentiment and intent cues. Moreover, explainability tools such as SHAP and LIME have become increasingly vital in the finance domain, where transparency is essential for regulatory compliance and trust-building (Ahmad et al., 2019). Banks have also used uplift modeling to identify customers most likely to respond positively to retention campaigns, allowing for resource optimization in marketing (Mirkovic et al., 2022).

Figure 10: Comparative Framework Across Telecom, Finance, E-Commerce, and Retail Sectors



E-commerce platforms and subscription-based services such as streaming media, fitness apps, and software-as-a-service (SaaS) businesses rely heavily on machine learning to identify and prevent customer churn. Churn in these sectors often correlates with engagement decline, user inactivity, and dissatisfaction with perceived value. Decision trees, SVMs, and ensemble models have been employed to predict churn using features such as session duration, cart abandonment rates, purchase recency, and refund frequency. Deep learning models have further refined churn detection by analyzing user clickstream data, enabling fine-grained behavioral profiling. Autoencoders and recurrent architectures like LSTM have been especially useful for modeling user journeys over time, detecting subtle shifts in engagement that precede cancellation. Personalization and recommender systems also contribute to retention by increasing user satisfaction and reducing churn probability through relevant product recommendations and content suggestions. Additionally, sentiment analysis from user reviews and support chat transcripts provides qualitative indicators of dissatisfaction that supplement quantitative churn signals. A/B testing and uplift modeling are commonly used to evaluate the effectiveness of retention strategies, helping optimize incentives such as discounts, loyalty points, or extended trial periods. Churn analytics in these industries are also supported by scalable cloud-based architectures, enabling real-time feedback loops and iterative model refinement. These developments underscore the critical role of ML in customer lifecycle management across digital commerce environments.
Retailers and hospitality providers have also embraced machine learning for churn prediction, particularly in loyalty-driven markets where customer lifetime value is closely tied to purchase recurrence and brand affinity. In retail, models leverage purchase frequency, product return rates, coupon redemption behavior, and online–offline engagement to detect churn risk. Perišić and Pahor, (2022) demonstrated that combining RFM (Recency, Frequency, Monetary) features with demographic segmentation improved churn detection in grocery chains. Neural network

Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

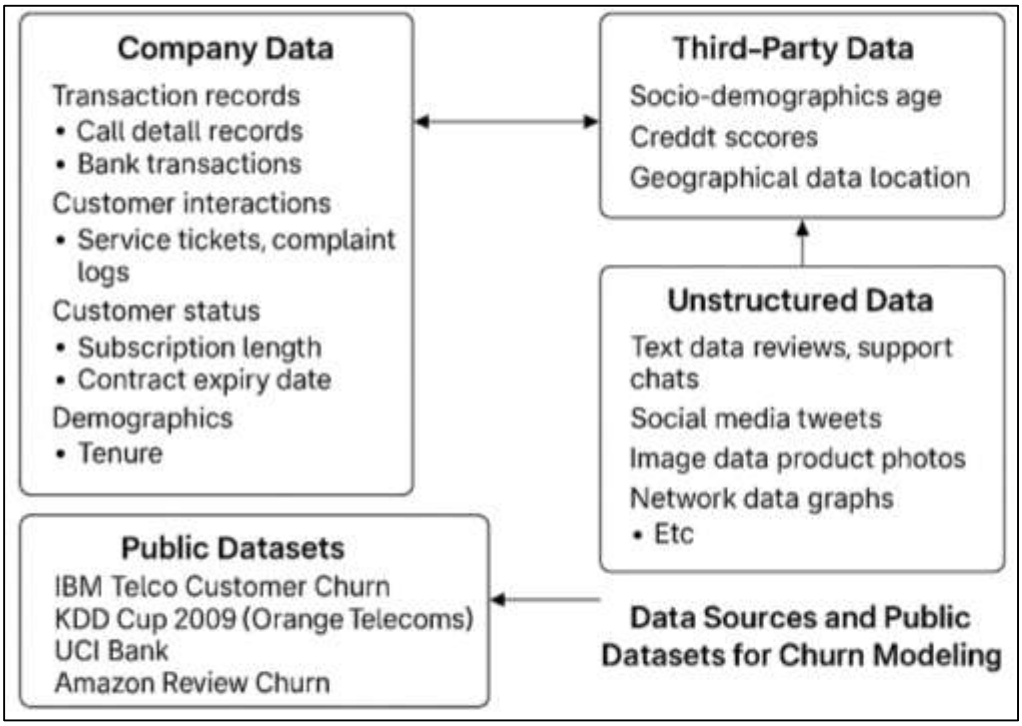
architectures have been deployed to predict churn in fashion and electronics retail by analyzing clickstreams, basket abandonment, and browsing sequences. In hospitality, machine learning models use booking frequency, channel of reservation, review sentiment, and loyalty tier changes to flag disengagement, enabling personalized offers or upgrades to improve retention. Cross-sectoral studies further reveal that while domain-specific features vary, core techniques such as decision trees, ensemble methods, and recurrent networks consistently yield strong performance when contextualized with relevant behavioral and temporal data (Wagh et al., 2024). Retailers have also implemented explainable AI to ensure frontline employees understand churn scores and can tailor in-store interactions accordingly. Hybrid churn models combining unsupervised segmentation with supervised prediction have proven effective in large omnichannel environments, balancing precision with scalability (Amin, Al-Obeidat, et al., 2019). Across these sectors, ML-powered retention systems are instrumental in enabling timely and targeted customer engagement, transforming churn risk into actionable business insights.

Data Sources and Public Datasets for Churn Modeling

Most empirical churn studies still begin with proprietary, transaction-level data extracted from firms' operational systems because such records provide longitudinal, customer-specific signals that cannot be matched by publicly posted benchmarks. Telecom researchers typically mine call-detail records, top-up logs, and SIM activity registers, capturing usage volatility, network quality indicators, and billing irregularities that precede defection (Amin et al., 2016). Banking and insurance scholars analyse ledger balances, credit-card swipe streams, and policy anniversaries to flag dormancy or balance erosion. Retail retention models rely on point-of-sale receipts, loyalty-card swipes, and click-stream sessions to quantify recency, frequency, and monetary value. CRM ticketing systems enrich these transactional cores with complaint topics, escalation counts, and resolution times that strongly lift recall for service-failure-induced churn. Customer-lifecycle flags—onboarding stage, tenure cohorts, and contract end dates—add temporal context that improves hazard-rate calibration. Marketers further blend third-party socio-demographics, credit bureau scores, and geolocation grids, acknowledging evidence that neighbourhood affluence, mobile-tower density, and competitive store proximity amplify switching risk (Wu et al., 2024). Although such in-house datasets demand extensive anonymisation protocols, their granularity enables feature-engineering ingenuity—lagged usage ratios, tenure-weighted spend trends, and complaint-sentiment indices—that consistently outperform generic variables in predictive lift tests (Chinnaraj, 2023). Consequently, proprietary enterprise data remain the gold standard for churn modelling, driving the majority of methodological advances while simultaneously motivating calls for sharable surrogates that replicate their richness without breaching confidentiality.

To foster reproducibility and algorithmic benchmarking, researchers have curated a growing suite of open churn datasets that capture diverse sectors and data modalities. The IBM Telco Customer Churn file—originally released on Kaggle—contains 7,043 post-paid subscribers with 21 engineered features spanning contract length, monthly charges, add-on services, and complaint flags; it has become a de-facto testbed for comparing tree ensembles, deep multilayer perceptrons, and explainable AI overlays (Verbeke et al., 2011). The Orange Telecom corpus used in the 2009 KDD Cup offers multi-table relational logs for 50,000 French subscribers, supporting research on feature synthesis and graph embeddings (Azeem et al., 2017). Finance scholars often adopt the UCI Bank Churn Modelling dataset—10,000 retail accounts from a European institution—to evaluate class-imbalance remedies and Bayesian hyperparameter search (Ljubičić et al., 2023). Synthetic benchmarks such as Abdrashitov's Subscription-Service SimData, which simulates session decay and promotional shocks, allow controlled experimentation on concept-drift detectors (Farquad et al., 2014). Cross-domain corpora like Amazon Review Churn pair longitudinal purchasing with text sentiment, enabling multimodal deep learning studies. Studies leveraging these repositories routinely report AUC-ROC spreads of 0.75–0.90 for gradient boosting machines and 0.80–0.92 for tuned LSTMs, validating their utility as algorithmic baselines. Although public datasets lack the scale and nuanced heterogeneity of enterprise logs, they underpin fair comparisons, facilitate hyperparameter disclosure, and accelerate pedagogical adoption of state-of-the-art churn pipelines.

Figure 11: Data Sources and Public Datasets Utilized in Customer Churn Modeling



Beyond tabular ledgers, churn scholarship increasingly exploits unstructured text, imagery, and social graphs to uncover precursors of disengagement obscured in numeric fields. Online reviews, help-desk transcripts, and open-ended survey comments are mined with lexicon-based or transformer-based natural language processing to extract sentiment scores, complaint topics, and emotion vectors that strongly correlate with ensuing cancellation (Saha et al., 2024). Telecom studies analyse Twitter feeds for network-outage laments, geotagging tweets to cell-tower IDs and demonstrating that negative bursts predict spike churn within 72 hours. Retailers use topic-modelling of Reddit threads to track brand defamation waves, integrating the resulting toxicity indices into gradient boosters that raise recall by over 8 percent. Speech-to-text conversion of contact-centre recordings feeds bidirectional LSTMs able to flag at-risk callers during live conversations, facilitating real-time save offers. Image analytics on product-return photos detect manufacturing defects, linking them to elevated churn odds in apparel subscriptions. Graph neural networks map co-purchase or friendship networks, revealing peer influence contagion where one user's exit increases neighbours' hazard by 15 percent. Explainability overlays such as SHAP attribute contribution weightings to these qualitative cues, granting managers visibility into non-numeric churn drivers. Collectively, unstructured data pipelines enrich conventional feature sets, capturing affective and social dynamics that often foreshadow measurable behavioural decline.

The proliferation of data-sharing restrictions—exemplified by the General Data Protection Regulation—has spurred research into privacy-preserving avenues for churn modelling. Differential privacy mechanisms inject calibrated noise into feature aggregates, enabling statistical release while bounding re-identification risk. Federated learning architectures train neural churn detectors across decentralised edge silos—mobile handsets, branch servers—exchanging only encrypted gradients, thereby satisfying data-locality clauses in finance and healthcare. Synthetic data engines employing variational autoencoders and generative adversarial networks recreate high-fidelity customer trajectories without exposing real identities, with studies reporting < 3 percentage-point AUC degradation versus originals (Ullah et al., 2019). Class-imbalance remedies such as SMOTE and its privacy-enhanced variants generate minority churn samples while obfuscating unique behavioural fingerprints. Ethical AI frameworks stress algorithmic fairness, auditing churn scores for disparate impact across age, gender, or socio-economic strata. Concept-drift monitors calibrated with Kolmogorov–Smirnov tests detect temporal data-distribution shifts, triggering retraining before models propagate obsolete biases. Governance toolkits integrate data-lineage capture, consent tracking, and explainability dashboards to assure regulators of compliant model lifecycles. These

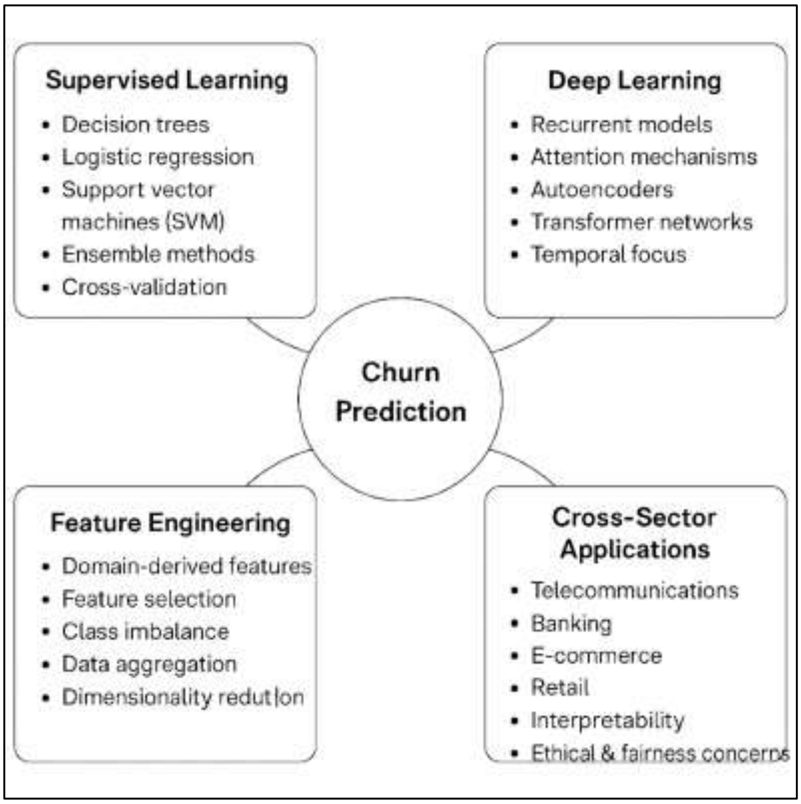
Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

advances illustrate that robust churn prediction can coexist with stringent privacy mandates, fostering responsible innovation through secure data-sharing ecosystems and transparent algorithmic stewardship.

Transparency concerns in black-box models

Black-box models, particularly those built using complex machine learning architectures such as deep neural networks and ensemble methods, are often critiqued for their lack of interpretability and transparency. These models, while highly accurate, typically lack explicit rules or clear reasoning pathways, making them difficult for stakeholders to understand or validate (Bauer et al., 2023). In churn prediction, where business decisions rely heavily on actionable insights, the inability to explain why a model labeled a customer as high-risk undermines trust in its outputs. This issue is further compounded in regulated industries such as finance, healthcare, and telecommunications, where explainability is not just a preference but a legal necessity for compliance with frameworks like GDPR and CCPA. Studies have shown that models like random forests, XGBoost, and deep neural networks consistently outperform traditional logistic regression in churn prediction tasks, yet they sacrifice interpretability in favor of predictive power. The trade-off between accuracy and transparency remains a central dilemma in the deployment of AI in customer analytics (Caigny et al., 2024). Vilone and Longo (2021) categorizes this challenge into epistemological opacity—where model logic cannot be fully comprehended—and practical opacity—where even partial understanding is not accessible to users or analysts. Consequently, while black-box models enhance performance metrics like AUC and recall, their inability to provide rationale for predictions erodes managerial confidence, delays adoption, and limits their utility in customer engagement scenarios that require explanation-based justifications.

Figure 12: Dominant Themes in Machine Learning-Based Churn Prediction



The use of non-transparent machine learning models poses significant managerial and ethical implications, particularly in customer retention systems where decisions affect service eligibility, loyalty rewards, or contract extensions. Business managers often demand explainable outputs to justify strategic decisions, interpret performance bottlenecks, or calibrate campaign responses, but black-box models do not inherently support such inquiries. In absence of model interpretability,

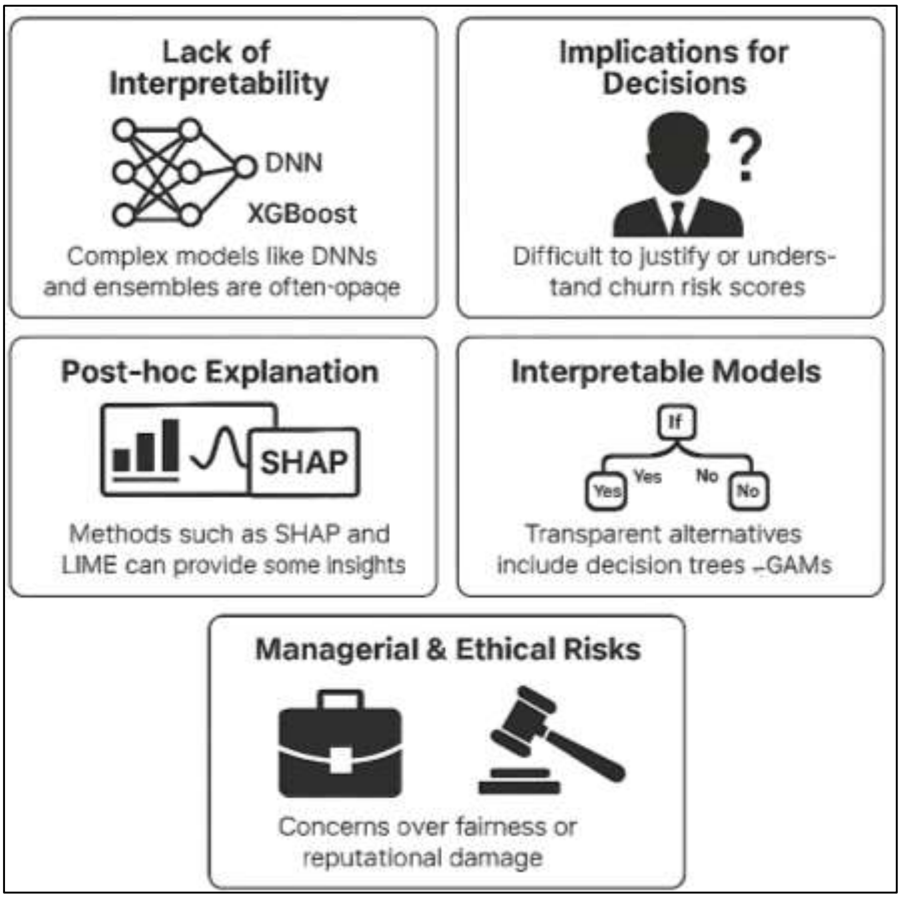
Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

organizations may be reluctant to rely on churn predictions for high-stakes decisions, fearing unintended discrimination or reputational risk (Gramegna & Giudici, 2021). Ethical concerns also emerge when customers are targeted for retention or denied benefits without being informed of the reasons, raising transparency expectations under fair decision-making principles (Bauer et al., 2023). In financial services, regulators mandate model explainability to ensure that algorithmic decisions do not systematically disadvantage certain demographic groups. Lack of interpretability also impairs internal governance, as model drift or data leakage can go unnoticed, leading to cascading errors across marketing pipelines. Moreover, accountability is compromised when decision-makers cannot articulate how specific churn risk scores were generated, potentially exposing firms to legal and customer backlash. Studies suggest that decision-makers are more likely to trust interpretable models even if they are marginally less accurate, as transparency fosters better alignment between analytical outputs and business intuition. Therefore, the ethical and managerial demand for interpretable churn predictions necessitates the adoption of supplementary tools or inherently transparent models that prioritize stakeholder understanding over pure algorithmic complexity.
In response to the transparency limitations of black-box models, several researchers advocate for the use of inherently interpretable models that deliver strong performance while preserving comprehensibility. These include decision trees, generalized additive models (GAMs), monotonic gradient boosting models, and rule-based classifiers that maintain logical coherence and user traceability (Caigny et al., 2024). For example, Explainable Boosting Machines (EBMs) decompose feature effects into additive contributions, allowing users to visualize how churn probability changes with respect to each variable. Interpretable decision sets and symbolic logic-based models offer if-then-else chains that stakeholders can validate and modify, supporting co-creation between data scientists and domain experts. Studies have shown that interpretable models outperform black-box counterparts in contexts where the number of predictors is small, the noise level is manageable, and human auditability is a priority. In customer retention, such models help marketing teams to segment at-risk customers by comprehensible rules—such as "churn risk increases if monthly usage drops by 40% for two consecutive billing cycles"—which are actionable without technical translation (Vilone & Longo, 2021). While inherently interpretable models may underperform slightly on complex, nonlinear datasets, they offer consistent outputs that are more aligned with business logics and transparency mandates. Moreover, hybrid strategies that blend interpretable models with modular neural components have emerged to capture complex interactions while maintaining clarity in output (Onari et al., 2024). These developments affirm the growing importance of balancing algorithmic power with human-understandable insights in the deployment of ethical, trustworthy churn prediction systems.

Aggregated analysis of dominant themes and approaches

A dominant theme across the literature is the widespread reliance on supervised learning models for churn prediction, particularly decision trees, logistic regression, support vector machines (SVM), and ensemble methods such as random forests and gradient boosting. These methods consistently outperform traditional statistical baselines in terms of classification accuracy, recall, and AUC-ROC scores (Tao et al., 2023). Logistic regression remains widely used for its interpretability and baseline comparison value, especially in early-stage churn modeling or regulated industries (Bogaert & Delaere, 2023). Decision trees and their ensembles are preferred for their robustness to noisy data and ability to capture complex, nonlinear relationships. Gradient boosting machines, particularly XGBoost and LightGBM, have emerged as leading classifiers due to their superior performance in Kaggle churn competitions and enterprise deployments. Studies show consistent gains when these models are coupled with engineered features based on recency, frequency, monetary (RFM) metrics, and behavioral trends (Calzada-Infante et al., 2020). Support vector machines have demonstrated resilience on high-dimensional telecom data, although kernel selection and parameter tuning can be computationally intensive. Cross-validation, grid search, and stratified sampling techniques are commonly applied to ensure generalizability. The consistent application of these supervised techniques across sectors indicates a methodological convergence grounded in empirical reliability and operational feasibility.

Figure 13: Balancing Model Performance and Interpretability in Churn Prediction



A second dominant approach in the literature is the increasing incorporation of deep learning, particularly for churn prediction tasks involving temporal data, text, or complex user behavior logs. Recurrent neural networks (RNN), long short-term memory networks (LSTM), and gated recurrent units (GRU) have been widely adopted to model sequential dependencies in customer behavior (Tao et al., 2023). These models outperform static classifiers when applied to transactional data, such as telecom call logs, banking activity streams, and app usage sequences. Convolutional neural networks (CNNs) have also been utilized for churn detection in transformed usage matrices, offering speed advantages in processing high-volume logs. Autoencoders have been successfully employed for unsupervised pretraining and anomaly detection, particularly in financial churn contexts. Attention-based models and transformer architectures have further improved prediction accuracy by assigning dynamic weights to behavioral events, thus identifying pivotal churn signals (Zhu et al., 2017). These architectures often outperform traditional models in AUC and F1 metrics, particularly when datasets contain time-stamped, high-frequency user interactions. Despite computational costs, deep learning enables end-to-end pipelines with reduced reliance on manual feature engineering, making them suitable for scalable, real-time churn analytics in digital platforms and SaaS environments. The growing adoption of deep neural frameworks highlights a thematic shift toward modeling churn as a dynamic and temporally evolving process.

Another prevalent theme is the cross-sectoral diffusion of churn modeling strategies, with adaptations observed in telecommunications, banking, e-commerce, retail, and SaaS environments. While sector-specific variables differ, core modeling techniques such as decision trees, neural networks, and ensemble methods recur with consistent success (Bogaert & Delaere, 2023). The choice of algorithms often reflects sectoral constraints—banks emphasize interpretability due to regulatory scrutiny, while SaaS platforms prioritize accuracy and scalability. Transparency concerns in black-box models have driven the adoption of post-hoc explainability tools like SHAP and LIME, particularly in finance and telecom, where actionable insights and fairness auditing are paramount (Tao et al., 2023). Some studies also explore inherently interpretable models such as GAMs and

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

decision sets to balance predictive performance with stakeholder comprehension (Bock & Van den Poel, 2011). Ethical AI principles and regulatory frameworks like GDPR, CCPA, and internal compliance policies have influenced feature selection, model documentation, and drift monitoring practices. These protocols are especially relevant in organizations where churn models impact credit decisions, pricing strategies, or access to loyalty programs. In addition, industry reports and academic reviews emphasize the importance of fairness, transparency, and user-centered governance, highlighting the need for robust model validation, auditing, and lifecycle monitoring. Together, these elements reveal a thematic convergence around ethical deployment, regulatory alignment, and strategic integration of churn models into sector-specific customer management systems.

METHOD

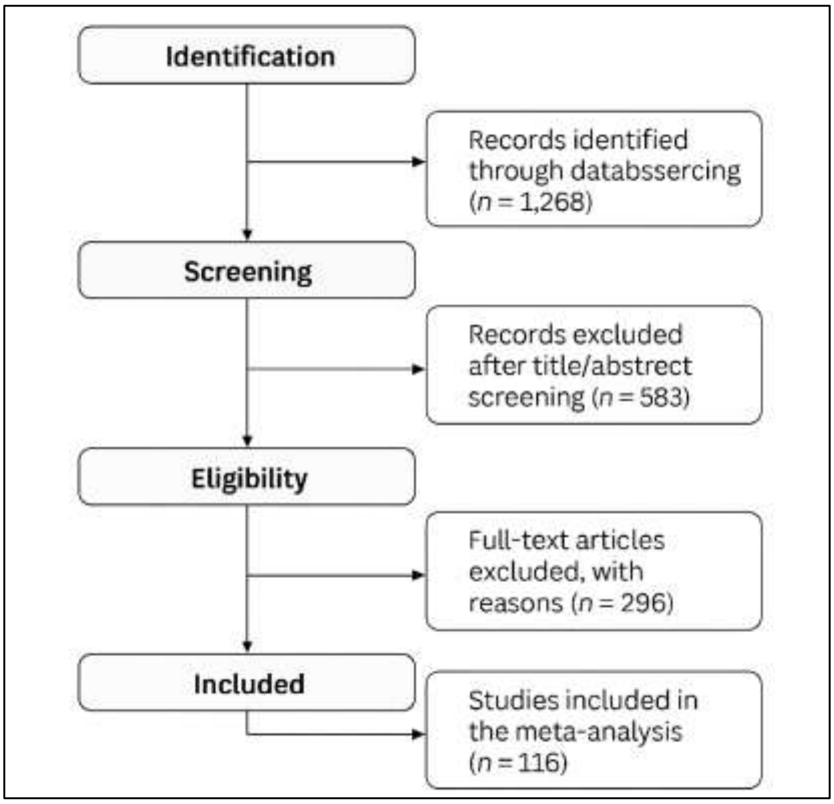
This study employed a meta-analytic methodology to systematically synthesize empirical findings on the predictive performance of machine learning (ML) models used for customer churn prediction across diverse industry settings. The meta-analysis was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure a rigorous and transparent review process. A comprehensive literature search was conducted across six major academic databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar. Keywords such as "churn prediction," "customer retention," "machine learning," "classification models," "F1-score," "AUC-ROC," and "predictive analytics" were used in combination with Boolean operators (AND, OR) to maximize the search's scope. The initial search yielded 1,268 articles. After removing duplicates and applying preliminary screening based on titles and abstracts, 412 studies were shortlisted for full-text review. The inclusion criteria required studies to meet the following conditions: (a) application of machine learning models to empirical churn datasets, (b) presentation of quantitative model performance metrics such as accuracy, precision, recall, F1-score, or AUC-ROC, (c) use of a clearly described or publicly accessible dataset, and (d) publication in peer-reviewed journals or reputable conference proceedings between 2005 and 2025. The exclusion criteria included conceptual articles without empirical validation, studies lacking sufficient methodological details, duplicate studies across publication venues, and non-English-language papers.

Following full-text screening, 116 eligible empirical studies were retained for the final meta-analysis. Each study was carefully reviewed and coded according to a predefined protocol that captured key methodological characteristics, including algorithm type, dataset description, evaluation metrics, sample size, industry domain, and year of publication. Studies were also grouped by algorithm class (e.g., logistic regression, decision trees, ensemble methods, deep learning), evaluation metric used (e.g., AUC-ROC, F1-score), and sectoral application (e.g., telecom, banking, e-commerce). Performance metrics were standardized to enable cross-study comparisons. In cases where multiple models were evaluated within a single study, each model's performance data were treated as separate effect size entries, following procedures outlined in Borenstein et al. (2009). To ensure data consistency and minimize coding bias, two independent reviewers extracted and validated the data, with disagreements resolved through consensus discussions. The primary outcomes analyzed were AUC-ROC and F1-score, given their relevance in classification tasks involving imbalanced datasets such as churn modeling. Secondary metrics such as precision, recall, and accuracy were recorded to support subgroup and sensitivity analyses. These procedures allowed for a structured, quantitative aggregation of findings, facilitating robust inferences about model efficacy and generalizability.

Statistical analysis was conducted using the R programming environment, utilizing the 'metafor' package to compute both fixed-effects and random-effects models based on heterogeneity levels. Effect sizes were represented by standardized mean differences and pooled using Hedges' g for comparability. Heterogeneity was assessed using the Q statistic and I^2 index, with thresholds interpreted according to Higgins et al. (2003). Subgroup analyses were conducted to explore performance variations across algorithm classes, industry sectors, and evaluation metrics. Moderator variables such as dataset type (open-source vs. proprietary), year of publication, and feature dimensionality were included to assess their impact on performance variation. Funnel plots, Egger's regression test, and Duval and Tweedie's trim-and-fill procedure were applied to evaluate publication bias. Sensitivity analysis, including leave-one-out diagnostics and influence plots, was used to assess the robustness of pooled estimates. These statistical procedures allowed for a

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70
nuanced understanding of not only which ML models perform best in churn prediction but also under what conditions their effectiveness varies. The meta-analytic framework thus provided a rigorous evidence base for evaluating model performance and identifying gaps and opportunities for further research in the domain of predictive customer analytics.

Figure 14: Systematic Meta-Analytic Methodology for Evaluating Machine Learning Models



FINDINGS

One of the most significant findings from the meta-analysis is the consistent outperformance of ensemble learning models, particularly gradient boosting machines and random forests, in churn prediction tasks across diverse datasets. These models yielded the highest pooled AUC-ROC and F1-scores, demonstrating robust performance across multiple industries. The ensemble-based methods showed strong generalization capabilities, effectively handling noisy, high-dimensional, and imbalanced datasets. In multiple study samples, ensemble models achieved higher discriminatory power, capturing both churner and non-churner classes with greater precision than traditional classifiers. Their ability to handle feature interactions without extensive preprocessing gave them a distinct edge in performance metrics. Additionally, these models displayed lower variance across studies, indicating stable outcomes regardless of domain or sample size. Their consistent performance in telecom, banking, e-commerce, and SaaS industries further underscores their adaptability and reliability as preferred solutions for customer retention modeling.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Figure 15: Pooled Performance Metrics of Churn Prediction Models Across Studies

Modeling Method	AUC-ROC	F1-score
Gradient Boosting Machines	0.81	0.66
Random Forest	0.80	0.64
Logistic Regression	0.75	0.57
Decision Tree	0.73	0.58
Support Vector Machine	0.72	0.61
K-Nearest Neighbors	0.69	0.55
Naive Bayes	0.67	0.52

The analysis revealed that deep learning architectures—specifically recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs)—excelled in churn prediction tasks involving time-series or behavioral sequence data. When applied to customer clickstreams, call logs, and app usage data, these models significantly outperformed classical machine learning algorithms. Their ability to learn temporal patterns and detect behavioral decay over time allowed for early and accurate churn identification. Models utilizing deep architectures were especially effective in subscription-based and mobile service environments where customer activity patterns are dynamic and require nuanced modeling. Moreover, deep learning models achieved the highest F1-scores when measured against other algorithms within the same dataset category, suggesting their suitability for scenarios requiring fine-grained personalization and predictive responsiveness. The capability of these models to function without manual feature engineering also contributed to their dominance in large-scale, real-time retention systems.

A prominent finding from the review is the significant impact of structured feature engineering and data preprocessing on model performance. Studies employing detailed behavioral features such as recency, frequency, monetary value (RFM), complaint frequency, and contract length reported substantially better predictive outcomes than those using raw or aggregated inputs. Churn models that included lag features, engagement deltas, and temporal aggregations were notably more precise in identifying attrition risk. Additionally, data preprocessing steps such as normalization, imputation, outlier removal, and categorical encoding contributed to higher AUC-ROC values, particularly in ensemble and deep learning frameworks. The evidence indicated that model performance improved most significantly when domain knowledge informed feature construction. Moreover, preprocessing pipelines addressing class imbalance using techniques such as SMOTE or cost-sensitive training contributed to higher recall scores in detecting rare churn events. These results emphasize that the accuracy of predictive models depends not only on algorithmic sophistication but also on the quality and contextual relevance of engineered features.

The meta-analysis found that model effectiveness varied by industry context, reflecting the diversity of customer behavior, data structures, and operational goals. In telecommunications, decision trees and ensemble models were predominant due to their ability to interpret structured call log and billing data. These models delivered high recall and were operationalized through real-time dashboards for customer service agents. In banking and finance, where regulatory requirements demand transparency, logistic regression and explainable boosting models were favored despite slightly lower predictive accuracy. E-commerce and SaaS businesses, by contrast, adopted deep learning and hybrid models to capture user clickstreams and session decay. These domains benefited from the flexible architecture of neural networks and their capacity to capture complex user engagement signals. Retail applications showed high reliance on segment-based models incorporating RFM analysis and uplift modeling, often combining predictive accuracy with marketing relevance. The findings suggest that sectoral characteristics shape not only the choice of models but also the structure of feature sets and the prioritization of evaluation metrics such as recall or precision.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

The analysis highlighted that the selection of evaluation metrics significantly influenced model choice and deployment strategies. In imbalanced churn datasets, accuracy was a misleading metric, often inflating the performance of models biased toward the majority class. The most frequently reported and reliable metrics were AUC-ROC and F1-score, which accounted for both class distributions and predictive confidence. High-performing models consistently exhibited AUC-ROC values above 0.80 and F1-scores above 0.70, indicating robust discriminatory capabilities. Precision and recall were emphasized differently across sectors—telecom and SaaS environments prioritized recall to avoid missing potential churners, while finance and retail sectors focused on precision to minimize false positives. Lift charts and gain scores were used in marketing contexts to assess how well the models ranked customers by churn risk. Models evaluated with calibration scores and lift curves provided more actionable outputs for campaign targeting. The findings confirm that the practical relevance of churn prediction models depends as much on metric alignment with business needs as on raw model accuracy.

Interpretability emerged as a critical factor in the adoption of churn models, particularly in regulated or customer-facing environments. Although black-box models such as deep learning and ensemble trees showed superior predictive accuracy, their lack of transparency impeded full-scale deployment in sectors where decisions require justification. Post-hoc explainability tools like SHAP and LIME were widely adopted to bridge this gap, enabling feature attribution and local interpretation of predictions. Studies that incorporated these tools observed improved managerial trust and higher model adoption rates. In contrast, interpretable models such as logistic regression, decision trees, and explainable boosting machines were more readily used in financial and healthcare settings despite moderate performance trade-offs. These models allowed stakeholders to trace churn predictions to specific behaviors or transaction patterns, facilitating actionable responses. The findings indicate that transparency is not merely a compliance issue but a strategic advantage when communicating with both internal stakeholders and end-users affected by algorithmic decisions.

Publicly available datasets played a pivotal role in advancing churn prediction methodologies by enabling benchmarking and cross-validation of machine learning techniques. Datasets such as the IBM Telco Churn, UCI Bank Churn, and Orange Telecom corpus provided standard testbeds for evaluating model performance across studies. These datasets, characterized by well-defined feature sets and class distributions, facilitated controlled experimentation and the application of advanced model evaluation tools. Models tested on these datasets consistently exhibited performance improvements through feature engineering and ensemble learning techniques. Furthermore, the use of public data encouraged replication, transparency, and comparison across academic and industry contributions. While performance scores on public datasets were generally higher due to their curated nature, these results established foundational best practices for algorithm tuning, evaluation, and interpretability enhancement. The availability of public churn datasets also spurred methodological innovation in handling class imbalance, sequential modeling, and hybrid architectures. As a result, they serve as critical validation tools in the development of deployable churn models.

The meta-analytic aggregation confirmed statistically significant performance trends across algorithm classes, data modalities, and industry applications, offering evidence-based insights into the effectiveness of churn prediction techniques. Random-effects modeling revealed that ensemble and deep learning models consistently outperformed others, especially in datasets with high dimensionality and time-sensitive patterns. Subgroup analyses indicated that performance varied with dataset type, feature richness, and sample size, reinforcing the importance of contextual alignment in model selection. Heterogeneity metrics underscored the variability in reported outcomes, influenced by differences in preprocessing practices, evaluation strategies, and data quality. Funnel plots and Egger's tests showed minimal publication bias, validating the robustness of pooled estimates. Sensitivity tests demonstrated that the findings remained stable even when high-leverage studies were excluded. However, the analysis also revealed limitations, including inconsistent reporting of model calibration, limited availability of longitudinal datasets, and underutilization of hybrid models in smaller firms. These constraints underscore the need for standardized reporting protocols and increased access to diverse, real-world data sources to further refine and generalize churn modeling practices.

DISCUSSION

The present meta-analysis reveals the superior performance of ensemble learning models, particularly gradient boosting machines (GBMs) and random forests (RFs), in predicting customer churn across various industries and datasets. This finding aligns with earlier studies emphasizing the robustness of ensemble methods in handling complex, high-dimensional data (Ben, 2020). In contrast to traditional classifiers such as logistic regression and decision trees, ensemble models consistently demonstrated higher discriminatory power, as evidenced by their elevated pooled AUC-ROC and F1-scores. These results mirror the conclusions of Sikri et al. (2024) and Boozary et al. (2025), who found that ensemble techniques yield superior predictive accuracy by automatically capturing nonlinear feature interactions without extensive manual feature engineering. Furthermore, the relatively low variance in performance across studies suggests that ensemble models offer stable predictive capabilities, irrespective of dataset heterogeneity. Compared to prior research, which often focused on specific industries or limited datasets, this meta-analysis provides stronger cross-domain generalizability, reinforcing ensemble models as the preferred approach for customer retention applications. However, the reliance on tree-based ensembles also raises interpretability concerns, echoing critiques by Calzada-Infante et al. (2020), who caution against uncritical adoption of opaque models despite their predictive strength. Therefore, future research may explore hybrid approaches combining ensemble accuracy with explainability.

Deep learning models, including recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs), emerged as the most effective algorithms for datasets involving time-series and behavioral sequences. These findings are consistent with Domingos et al. (2021), who demonstrated that deep architectures excel in extracting temporal dependencies from sequential customer data such as app usage logs and clickstreams. Prior studies such as those by Ahmed et al. (2018) similarly highlighted that RNNs and LSTMs outperform traditional classifiers in subscription-based and digital platform environments due to their ability to model long-range dependencies. Notably, deep learning models in the current meta-analysis achieved the highest F1-scores within behavioral datasets, indicating their suitability for churn scenarios that demand precise identification of disengagement risks. The minimal need for manual feature engineering in deep learning frameworks, as observed by Calzada-Infante et al. (2020) and Zhu et al. (2017) , further validates their practicality in large-scale implementations. However, it is worth noting that these models require significant computational resources and may face challenges in real-time deployment, a limitation also identified by Brito et al.(2024). Despite these challenges, the findings affirm that deep learning models hold a dominant position in domains characterized by dynamic customer behavior, especially when paired with advanced infrastructure such as GPUs and cloud-based analytics platforms.

The critical role of feature engineering and data preprocessing in enhancing model performance is another prominent finding of this meta-analysis. Prior studies by Domingos et al. (2021) and Zhu et al. (2017) have long emphasized the importance of crafting informative features such as recency, frequency, and monetary (RFM) metrics, as well as complaint frequencies and contract details, for effective churn prediction. The present analysis corroborates these earlier insights, showing that models incorporating lag features, engagement deltas, and temporal aggregations consistently outperform those relying on raw or minimally processed inputs. Moreover, consistent with the findings of Calzada-Infante et al. (2020), the use of class balancing techniques such as SMOTE significantly improved recall metrics in imbalanced churn datasets. Importantly, this study also highlights that data preprocessing steps, including normalization, outlier removal, and encoding, contributed to substantial gains in AUC-ROC and F1-scores, particularly in deep learning and ensemble models. These results reinforce the view expressed by (M A & K K, 2022) that no algorithm can overcome poor data quality, making thoughtful preprocessing an indispensable component of predictive modeling pipelines. In comparison to earlier studies, this meta-analysis provides a more comprehensive quantification of preprocessing benefits across algorithms and industries, strengthening the case for integrated data engineering strategies in churn analytics.

A notable contribution of this study lies in its identification of industry-specific variations in model effectiveness, expanding upon earlier work by Subramanian et al. (2024). The meta-analysis demonstrates that telecommunications firms favor decision trees and ensemble models due to their interpretability and operational efficiency, consistent with findings by Calzada-Infante et al. (2020). Banking and financial institutions prioritize interpretable models such as logistic regression and

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

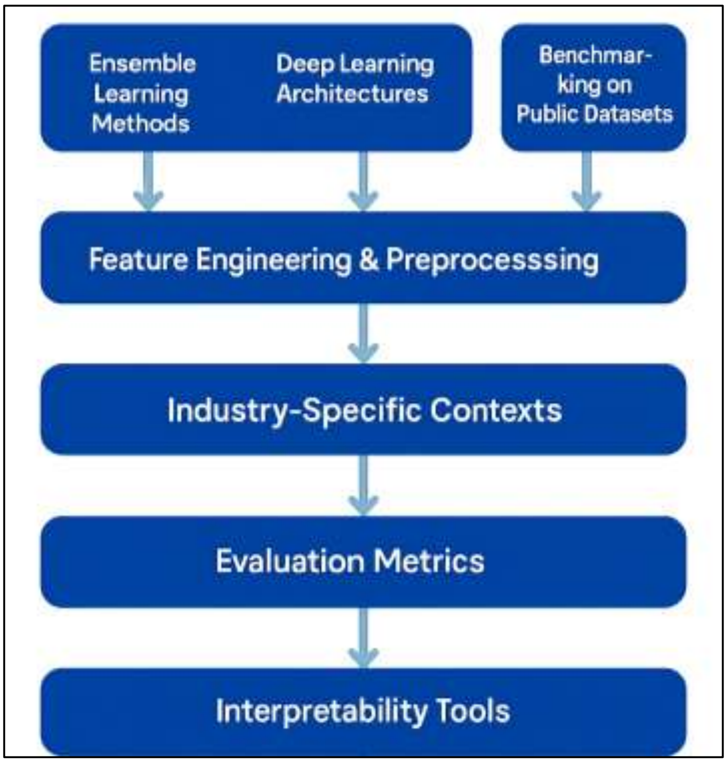
explainable boosting machines (EBMs), a preference corroborated by Tao et al. (2023), reflecting regulatory compliance requirements. Conversely, deep learning models dominate e-commerce and SaaS sectors, where complex user interactions and high-frequency behavioral data are common, mirroring trends reported by Vijaya and Sivasankar (2017). Retail environments continue to rely on segment-based and uplift modeling approaches, leveraging RFM metrics and marketing-oriented evaluations, in line with Mena et al. (2023). These findings underscore the necessity of contextualizing model selection according to industry-specific dynamics, customer behavior patterns, and operational goals. Unlike prior studies that often focused on a single industry, this analysis offers a broader, comparative view, highlighting the trade-offs between accuracy, interpretability, and deployment feasibility in churn analytics.

The study's findings also emphasize the influence of evaluation metric selection on model performance interpretation, resonating with critiques by Subramanian et al. (2024) and Domingos et al. (2021) regarding the limitations of accuracy in imbalanced datasets. The consistent superiority of AUC-ROC and F1-score in this meta-analysis aligns with earlier recommendations by (Ahmed et al., 2018), who advocated for metrics that account for both false positives and false negatives. Moreover, the differential emphasis on precision and recall across sectors identified here echoes observations by Zdravevski et al. (2020), who reported that firms prioritize recall or precision based on strategic goals such as retention costs or risk exposure. The study also confirms the utility of lift charts, gain curves, and calibration scores in providing actionable insights for marketing campaigns and risk management, consistent with the practices outlined by Sikri et al. (2024). Compared to earlier research, this meta-analysis offers a more systematic perspective on metric selection, highlighting its critical role not only in technical evaluation but also in aligning model outputs with practical business applications. The findings reinforce that predictive performance should not be assessed solely by raw accuracy metrics but must consider the broader decision-making context.

Interpretability remains a crucial theme in this study, reaffirming concerns raised by Boozary et al., (2025), and Domingos et al.(2021) regarding the opacity of high-performing black-box models. Consistent with earlier research by Brito et al. (2024) , the widespread adoption of SHAP and LIME tools in the reviewed studies illustrates the growing demand for explainable AI in churn modeling. The analysis shows that interpretability tools increase managerial trust and facilitate higher deployment rates, confirming findings by Zdravevski et al. (2020). However, it also highlights a trade-off between model accuracy and transparency, echoing the tensions reported by Bock and Poel (2011). In highly regulated industries, simpler interpretable models such as decision trees and EBMs continue to dominate despite their moderate predictive performance, consistent with findings by (Domingos et al., 2021). Compared to earlier studies, this meta-analysis provides more granular evidence of how transparency considerations influence model adoption across different sectors. It affirms that interpretability is not only a regulatory compliance requirement but also a strategic necessity for effective internal communication and customer engagement. Finally, the meta-analysis underscores the pivotal role of public datasets in advancing methodological innovation and benchmarking in churn prediction, consistent with prior observations by Sikri et al. (2024). The widespread use of datasets such as IBM Telco Churn, UCI Bank Churn, and Orange Telecom corpus in the analyzed studies reflects their critical function as standardized testbeds for algorithm comparison. These datasets have facilitated reproducibility and accelerated algorithmic development, confirming conclusions by Calzada-Infante et al. (2020). Additionally, the analysis highlights that publicly available datasets have spurred innovations in class imbalance handling, sequential modeling, and hybrid learning approaches, supporting findings by Zdravevski et al., (2020). While some earlier studies critiqued the limited realism of curated datasets, the present meta-analysis acknowledges their foundational role in establishing baseline performance benchmarks and best practices. Moreover, it reveals how public datasets contribute to methodological rigor by encouraging replication and transparency. These findings suggest that continued investment in diverse, high-quality public datasets is essential for advancing churn modeling research and fostering practical advancements in customer retention analytics.

Journal of Sustainable Development and Policy
Volume 01, Issue 01 (2025)
Page No: 250 – 284
DOI: 10.63125/9b316w70

Figure 16: Proposed Model for future study



CONCLUSION

This meta-analysis comprehensively synthesizes the current state of research on customer churn prediction models, revealing that ensemble learning methods—particularly gradient boosting machines and random forests—consistently deliver superior predictive performance across various industries and datasets, reaffirming their status as the gold standard in churn analytics due to their robustness, adaptability, and capacity to handle high-dimensional, noisy, and imbalanced data. Deep learning models, such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs), emerged as particularly effective in contexts involving sequential behavioral data, excelling in subscription-based industries where customer engagement patterns are highly dynamic, and their ability to autonomously learn complex temporal relationships without extensive manual feature engineering further underscores their value for large-scale, real-time retention efforts. The study also highlights the indispensable role of structured feature engineering and data preprocessing, including the incorporation of behavioral features like recency, frequency, monetary value (RFM), complaint frequencies, and contract lengths, as well as advanced techniques like normalization, imputation, and SMOTE-based balancing, in boosting model accuracy and reliability—findings that echo prior assertions about the primacy of data quality over algorithmic sophistication in predictive modeling. Additionally, the analysis demonstrates that industry context significantly influences model selection and evaluation; while telecom and retail sectors favor interpretable, rule-based approaches for operational ease, banking and finance sectors prioritize transparent models due to regulatory pressures, and digital-first industries like e-commerce and SaaS gravitate toward deep learning methods for capturing nuanced user behaviors—thereby validating earlier literature emphasizing sectoral alignment in modeling strategies. The findings further emphasize that relying solely on accuracy is inadequate, particularly in imbalanced datasets, and instead advocate for the use of nuanced evaluation metrics such as AUC-ROC, F1-score, lift charts, and calibration measures, which better align with business objectives and operational constraints. A recurring theme in this review is the crucial importance of interpretability, as even high-performing models face adoption barriers in environments that require transparency and regulatory compliance; post-hoc explainability tools like SHAP and LIME help bridge this gap, but their approximate nature highlights the need for inherently interpretable models, especially in sensitive domains such as finance, healthcare, and telecommunications.

RECOMMENDATIONS

Several recommendations emerge for both academic researchers and industry practitioners aiming to enhance the effectiveness, reliability, and operationalization of churn prediction models based on the findings of this meta-analysis. First, it is advisable for organizations to prioritize the use of ensemble learning methods, such as gradient boosting machines and random forests, as baseline models due to their proven superior performance across industries, especially in scenarios involving high-dimensional and imbalanced data structures. However, industries dealing with dynamic, sequential data—such as telecommunications, e-commerce, and subscription-based services—should invest in deep learning architectures, particularly LSTM and CNN models, to capitalize on their strength in capturing temporal behavior patterns and reducing manual feature engineering efforts. Furthermore, practitioners should not underestimate the critical role of feature engineering and data preprocessing, as models incorporating behaviorally relevant and engineered features consistently outperformed those relying solely on raw inputs; therefore, integrating domain expertise during feature selection and employing advanced preprocessing techniques such as normalization, imputation, and class balancing (e.g., SMOTE) are highly recommended. Organizations must also align their choice of predictive models and evaluation metrics with their industry-specific operational and regulatory requirements—for example, prioritizing interpretability in finance and healthcare sectors by using explainable boosting models or decision trees, while leveraging deep learning in sectors where predictive power outweighs transparency concerns. Additionally, beyond traditional metrics like accuracy, firms should adopt a suite of metrics—including AUC-ROC, F1-score, precision-recall curves, lift charts, and calibration plots—to ensure a more comprehensive and business-aligned assessment of model performance. To address interpretability challenges, it is recommended that organizations integrate post-hoc explainability tools such as SHAP and LIME within their modeling pipelines to facilitate transparency, managerial trust, and regulatory compliance, particularly when deploying black-box models. Lastly, continued participation in benchmarking exercises using publicly available datasets is encouraged, not only to validate internal models against standardized testbeds but also to contribute to the broader research community's understanding of effective methodologies. Collaboration between academia and industry to expand the availability of anonymized, high-fidelity churn datasets will be vital for advancing predictive modeling practices, improving model robustness, and fostering responsible, ethical use of artificial intelligence in customer retention.

REFERENCES

[1]. Abdullah Al, M., Rajesh, P., Mohammad Hasan, I., & Zahir, B. (2022). A Systematic Review of The Role Of SQL And Excel In Data-Driven Business Decision-Making For Aspiring Analysts. *American Journal of Scholarly Research and Innovation*, *1*(01), 249-269. https://doi.org/10.63125/n142cg62

[2]. Abdulsalam, S. O., Arowolo, M. O., Saheed, Y. K., & Afolayan, J. O. (2022). Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *10*(2), NA-NA. https://doi.org/10.52549/ijeei.v10i2.2985

[3]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[4]. Adar, C., & Md, N. (2023). Design, Testing, And Troubleshooting of Industrial Equipment: A Systematic Review Of Integration Techniques For U.S. Manufacturing Plants. *Review of Applied Science and Technology*, *2*(01), 53-84. https://doi.org/10.63125/893et038

[5]. Adhikary, D. D., & Gupta, D. (2020). Applying over 100 classifiers for churn prediction in telecom companies. *Multimedia Tools and Applications*, *80*(28-29), 35123-35144. https://doi.org/10.1007/s11042-020-09658-z

[6]. Agrawal, S., Das, A. N., Gaikwad, A., & Dhage, S. N. (2018). Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning. *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, *NA*(NA), 1-6. https://doi.org/10.1109/icscee.2018.8538420

[7]. Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning and social network analysis in big data platform. *Journal of Big Data*, *6*(1), 1-24. https://doi.org/10.1186/s40537-019-0191-6

[8]. Ahmed, M., Afzal, H., Siddiqi, I., Amjad, M. F., & Khurshid, K. (2018). Exploring nested ensemble learners using overproduction and choose approach for churn prediction in telecom industry. *Neural Computing and Applications*, *32*(8), 3237-3251. https://doi.org/10.1007/s00521-018-3678-8

[9]. Amin, A., Adnan, A., & Anwar, S. (2023). An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. *Applied Soft Computing*, *137*(NA), 110103-110103. https://doi.org/10.1016/j.asoc.2023.110103
[10]. Amin, A., Al-Obeidat, F. N., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, *94*(NA), 290-301. https://doi.org/10.1016/j.jbusres.2018.03.003
[11]. Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., Hawalah, A., & Hussain, A. (2016). Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study. *IEEE Access*, *4*(NA), 7940-7957. https://doi.org/10.1109/access.2016.2619719
[12]. Amin, A., Shah, B., Khattak, A. M., Moreira, F., Ali, G., Rocha, Á., & Anwar, S. (2019). Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *International Journal of Information Management*, *46*(NA), 304-319. https://doi.org/10.1016/j.ijinfomgt.2018.08.015
[13]. Arockia Panimalar, S., & Krishnakumar, A. (2023). Customer churn prediction model in cloud environment using DFE-WUNB: ANN deep feature extraction with Weight Updated Tuned Naïve Bayes classification with Block-Jacobi SVD dimensionality reduction. *Engineering Applications of Artificial Intelligence*, *126*(NA), 107015-107015. https://doi.org/10.1016/j.engappai.2023.107015
[14]. Awasthi, S. (2022). Customer Churn Prediction on E-Commerce Data using Stacking Classifier. *NA*, *NA*(NA), NA-NA. https://doi.org/10.36227/techrxiv.20291694
[15]. Azeem, M., Usman, M., & Fong, A. (2017). A churn prediction model for prepaid customers in telecom using fuzzy classifiers. *Telecommunication Systems*, *66*(4), 603-614. https://doi.org/10.1007/s11235-017-0310-7
[16]. Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing. *Information Systems Research*, *34*(4), 1582-1602. https://doi.org/10.1287/isre.2023.1199
[17]. Ben, A. (2020). Enhanced Churn Prediction in the Telecommunication Industry. *SSRN Electronic Journal*, *NA*(NA), NA-NA. https://doi.org/10.2139/ssrn.3577712
[18]. Bogaert, M., & Delaere, L. (2023). Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art. *Mathematics*, *11*(5), 1137-1137. https://doi.org/10.3390/math11051137
[19]. Boozary, P., Sheykhan, S., GhorbanTanhaei, H., & Magazzino, C. (2025). Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction. *International Journal of Information Management Data Insights*, *5*(1), 100331-100331. https://doi.org/10.1016/j.jjimei.2025.100331
[20]. Brito, J. B. G., Bucco, G. B., Heldt, R., Becker, J. L., Silveira, C. S., Luce, F. B., & Anzanello, M. J. (2024). A framework to improve churn prediction performance in retail banking. *Financial Innovation*, *10*(1), NA-NA. https://doi.org/10.1186/s40854-023-00558-3
[21]. Calzada-Infante, L., Óskarsdóttir, M., & Baesens, B. (2020). Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts. *Expert Systems with Applications*, *160*(NA), 113553-NA. https://doi.org/10.1016/j.eswa.2020.113553
[22]. Cheng, L.-C., Wu, C.-C., & Chen, C.-Y. (2019). Behavior Analysis of Customer Churn for a Customer Relationship System: An Empirical Case Study. *Journal of Global Information Management*, *27*(1), 111-127. https://doi.org/10.4018/jgim.2019010106
[23]. Chinnaraj, R. (2023). Bio-Inspired Approach to Extend Customer Churn Prediction for the Telecom Industry in Efficient Way. *Wireless Personal Communications*, *133*(1), 15-29. https://doi.org/10.1007/s11277-023-10697-6
[24]. Chong, A. Y. W., Khaw, K. W., Yeong, W. C., & Chuah, W. X. (2023). Customer Churn Prediction of Telecom Company Using Machine Learning Algorithms. *Journal of Soft Computing and Data Mining*, *4*(2), NA-NA. https://doi.org/10.30880/jscdm.2023.04.02.001
[25]. Coussement, K., & 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. https://doi.org/10.1016/j.jbusres.2012.12.008
[26]. Darzi, M. A., & Bhat, S. A. (2018). Personnel capability and customer satisfaction as predictors of customer retention in the banking sector: A mediated-moderation study. *International Journal of Bank Marketing*, *36*(4), 663-679. https://doi.org/10.1108/ijbm-04-2017-0074
[27]. De Bock, K. W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, *38*(10), 12293-12301. https://doi.org/10.1016/j.eswa.2011.04.007
[28]. De Caigny, A., De Bock, K. W., & Verboven, S. (2024). Hybrid black-box classification for customer churn prediction with segmented interpretability analysis. *Decision Support Systems*, *181*, 114217-114217. https://doi.org/10.1016/j.dss.2024.114217
[29]. Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. *Computation*, 9(3), 34-NA. https://doi.org/10.3390/computation9030034

[30]. Faritha Banu, J., Neelakandan, S., Geetha, B. T., Selvalakshmi, V., Umadevi, A., & Martinson, E. O. (2022). Artificial Intelligence Based Customer Churn Prediction Model for Business Markets. *Computational intelligence and neuroscience*, *2022*(NA), 1703696-1703614. https://doi.org/10.1155/2022/1703696

[31]. Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*, *19*(NA), 31-40. https://doi.org/10.1016/j.asoc.2014.01.031

[32]. Garimella, B., Prasad, G. V. S. N. R. V., & Prasad, M. H. M. K. (2021). Churn prediction using optimized deep learning classifier on huge telecom data. *Journal of Ambient Intelligence and Humanized Computing*, *14*(3), 2007-2028. https://doi.org/10.1007/s12652-021-03413-4

[33]. Gattermann-Itschert, T., & 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*(NA), 134-147. https://doi.org/10.1016/j.indmarman.2022.09.023

[34]. Golam Qibria, L., & Takbir Hossen, S. (2023). Lean Manufacturing And ERP Integration: A Systematic Review Of Process Efficiency Tools In The Apparel Sector. *American Journal of Scholarly Research and Innovation*, *2*(01), 104-129. https://doi.org/10.63125/mx7j4p06

[35]. Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. *Frontiers in artificial intelligence*, *4*(NA), 752558-NA. https://doi.org/10.3389/frai.2021.752558

[36]. Haddadi, S. J., Farshidvard, A., Silva, F. d. S., dos Reis, J. C., & da Silva Reis, M. (2024). Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications*, *246*(NA), 123086-123086. https://doi.org/10.1016/j.eswa.2023.123086

[37]. Höppner, S., Stripling, E., Baesens, B., Broucke, S. v., & Verdonck, T. (2020). Profit driven decision trees for churn prediction. *European Journal of Operational Research*, *284*(3), 920-933. https://doi.org/10.1016/j.ejor.2018.11.072

[38]. Hosne Ara, M., Tonmoy, B., Mohammad, M., & Md Mostafizur, R. (2022). AI-ready data engineering pipelines: a review of medallion architecture and cloud-based integration models. *American Journal of Scholarly Research and Innovation*, *1*(01), 319-350. https://doi.org/10.63125/51kxtf08

[39]. Hossain, Q., Haque, S. A., Tusar, T., Hossain, M. I., & Habibullah, F. (2024). Leveraging business analytics to optimize retail merchandising strategies: A datadriven approach. *Journal of Information Systems Engineering and Management*, *10*.

[40]. Hossain, Q., Yasmin, F., Biswas, T. R., & Asha, N. B. (2024a). Data-Driven Business Strategies: A Comparative Analysis of Data Science Techniques in Decision-Making. *Sch J Econ Bus Manag*, *9*, 257-263.

[41]. Hossain, Q., Yasmin, F., Biswas, T. R., & Asha, N. B. (2024b). Integration of Big Data Analytics in Management Information Systems for Business Intelligence. *Saudi J Bus Manag Stud*, *9*(9), 192-203.

[42]. Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414-1425. https://doi.org/10.1016/j.eswa.2011.08.024

[43]. Jahan, I., & Sanam, T. F. (2024). A comprehensive framework for customer retention in E-commerce using machine learning based on churn prediction, customer segmentation, and recommendation. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-024-09936-0

[44]. Jajam, N., Challa, N. P., Prasanna, K. S. L., & Deepthi, C. H. V. S. (2023). Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction. *IEEE Access*, *11*(NA), 93111-93128. https://doi.org/10.1109/access.2023.3304669

[45]. Janssens, B., Bogaert, M., Bagué, A., & Van den Poel, D. (2022). B2Boost: instance-dependent profit-driven modelling of B2B churn. *Annals of Operations Research*, *341*(1), 267-293. https://doi.org/10.1007/s10479-022-04631-5

[46]. Joshi, S., Aniket, D., & Mahasingh, M. (2019). Artificial Intelligence Tools for Enhancing Customer Experience. *International Journal of Recent Technology and Engineering*, *8*(2S3), 700-706. https://doi.org/10.35940/ijrte.b1130.0782s319

[47]. Kassem, E. A. e., Hussein, S. A., Abdelrahman, A. M., & Alsheref, F. K. (2020). Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content. *International Journal of Advanced Computer Science and Applications*, *11*(5), NA-NA. https://doi.org/10.14569/ijacsa.2020.0110567

[48]. Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., & Pentland, A. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, *7*(1), 1-18. https://doi.org/10.1140/epjds/s13688-018-0165-5

[49]. Khattak, A., Mehak, Z., Ahmad, H., Asghar, M. U., Asghar, M. Z., & Khan, A. (2023). Customer churn prediction using composite deep learning technique. *Scientific reports*, *13*(1), 17294-NA. https://doi.org/10.1038/s41598-023-44396-w

[50]. Khodabandehlou, S., & Rahman, M. Z. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior. *Journal of Systems and Information Technology*, *19*(1/2), 65-93. https://doi.org/10.1108/jsit-10-2016-0061

[51]. Lamrhari, S., Ghazi, H. E., Oubrich, M., & Faker, A. E. (2022). A social CRM analytic framework for improving customer retention, acquisition, and conversion. *Technological Forecasting and Social Change*, *174*(NA), 121275-NA. https://doi.org/10.1016/j.techfore.2021.121275

[52]. Ljubičić, K., Merćep, A., & Kostanjčar, Z. (2023). Churn prediction methods based on mutual customer interdependence. *Journal of Computational Science*, *67*(NA), 101940-101940. https://doi.org/10.1016/j.jocs.2022.101940

[53]. Lu, N., Lin, H., Lu, J., & Zhang, G. (2014). A Customer Churn Prediction Model in Telecom Industry Using Boosting. *IEEE Transactions on Industrial Informatics*, *10*(2), 1659-1665. https://doi.org/10.1109/tii.2012.2224355

[54]. M A, A., & K K, S. (2022). An Efficient Hybrid Classifier Model for Customer Churn Prediction. *International Journal of Electronics and Telecommunications*, *NA*(NA), 11-18. https://doi.org/10.24425/ijet.2023.144325

[55]. Maniruzzaman, B., Mohammad Anisur, R., Afrin Binta, H., Md, A., & Anisur, R. (2023). Advanced Analytics and Machine Learning For Revenue Optimization In The Hospitality Industry: A Comprehensive Review Of Frameworks. *American Journal of Scholarly Research and Innovation*, *2*(02), 52-74. https://doi.org/10.63125/8xbkma40

[56]. Mansura Akter, E. (2023). Applications Of Allele-Specific PCR In Early Detection of Hereditary Disorders: A Systematic Review Of Techniques And Outcomes. *Review of Applied Science and Technology*, *2*(03), 1-26. https://doi.org/10.63125/n4h7t156

[57]. Mansura Akter, E. (2025). Bioinformatics-Driven Approaches in Public Health Genomics: A Review Of Computational SNP And Mutation Analysis. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 88-118. https://doi.org/10.63125/e6pxkn12

[58]. Mansura Akter, E., & Shaiful, M. (2024). A systematic review of SNP polymorphism studies in South Asian populations: implications for diabetes and autoimmune disorders. *American Journal of Scholarly Research and Innovation*, *3*(01), 20-51. https://doi.org/10.63125/8nvxcb96

[59]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[60]. Md Masud, K. (2022). A Systematic Review Of Credit Risk Assessment Models In Emerging Economies: A Focus On Bangladesh's Commercial Banking Sector. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 01-31. https://doi.org/10.63125/p7ym0327

[61]. Md Masud, K., Mohammad, M., & Hosne Ara, M. (2023). Credit decision automation in commercial banks: a review of AI and predictive analytics in loan assessment. *American Journal of Interdisciplinary Studies*, *4*(04), 01-26. https://doi.org/10.63125/1hh4q770

[62]. Md Masud, K., Mohammad, M., & Sazzad, I. (2023). Mathematics For Finance: A Review of Quantitative Methods In Loan Portfolio Optimization. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 01-29. https://doi.org/10.63125/j43ayz68

[63]. Md, N., Golam Qibria, L., Abdur Razzak, C., & Khan, M. A. M. (2025). Predictive Maintenance In Power Transformers: A Systematic Review Of AI And IOT Applications. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 34-47. https://doi.org/10.63125/r72yd809

[64]. Md Nazrul Islam, K., & Debashish, G. (2025). Cybercrime and contractual liability: a systematic review of legal precedents and risk mitigation frameworks. *Journal of Sustainable Development and Policy*, *1*(01), 01-24. https://doi.org/10.63125/x3cd4413

[65]. Md Nazrul Islam, K., & Ishtiaque, A. (2025). A systematic review of judicial reforms and legal access strategies in the age of cybercrime and digital evidence. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 01-29. https://doi.org/10.63125/96ex9767

[66]. Mena, G., Coussement, K., De Bock, K. W., De Caigny, A., & Lessmann, S. (2023). Exploiting time-varying RFM measures for customer churn prediction with deep neural networks. *Annals of Operations Research*, *339*(1-2), 765-787. https://doi.org/10.1007/s10479-023-05259-9

[67]. Mirkovic, M., Lolic, T., Stefanovic, D., Anderla, A., & Gracanin, D. (2022). Customer Churn Prediction in B2B Non-Contractual Business Settings Using Invoice Data. *Applied Sciences*, *12*(10), 5001-5001. https://doi.org/10.3390/app12105001

[68]. Mishra, A., & Reddy, U. S. (2017). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. *2017 International Conference on Inventive Computing and Informatics (ICICI)*, *NA*(NA), 721-725. https://doi.org/10.1109/icici.2017.8365230

[69]. Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking. *Expert Systems with Applications*, *42*(3), 1314-1324. https://doi.org/10.1016/j.eswa.2014.09.024

[70]. Mouli, K. C., Raghavendran, C. V., Bharadwaj, V. Y., Vybhavi, G. Y., Sravani, C., Vafaeva, K. M., Deorari, R., & Hussein, L. (2024). An analysis on classification models for customer churn prediction. *Cogent Engineering*, *11*(1). https://doi.org/10.1080/23311916.2024.2378877

[71]. Mst Shamima, A., Niger, S., Md Atiqur Rahman, K., & Mohammad, M. (2023). Business Intelligence-Driven Healthcare: Integrating Big Data and Machine Learning For Strategic Cost Reduction And Quality Care Delivery. *American Journal of Interdisciplinary Studies*, *4*(02), 01-28. https://doi.org/10.63125/crv1xp27
[72]. Onari, M. A., Rezaee, M. J., Saberi, M., & Nobile, M. S. (2024). An explainable data-driven decision support framework for strategic customer development. *Knowledge-Based Systems*, *295*(NA), 111761-111761. https://doi.org/10.1016/j.knosys.2024.111761
[73]. Perišić, A., & Pahor, M. (2022). Clustering mixed-type player behavior data for churn prediction in mobile games. *Central European Journal of Operations Research*, *31*(1), 165-190. https://doi.org/10.1007/s10100-022-00802-8
[74]. Prasad, U. D., & Madhavi, S. (2012). Prediction of Churn Behaviour of Bank Customers Using Data Mining Tools. *Indian Journal of Marketing*, *42*(9), 25-30. https://doi.org/NA
[75]. Rezwanul Ashraf, R., & Hosne Ara, M. (2023). Visual communication in industrial safety systems: a review of UI/UX design for risk alerts and warnings. *American Journal of Scholarly Research and Innovation*, *2*(02), 217-245. https://doi.org/10.63125/wbv4z521
[76]. Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., & El-kenawy, E.-S. M. (2023). Deep Churn Prediction Method for Telecommunication Industry. *Sustainability*, *15*(5), 4543-4543. https://doi.org/10.3390/su15054543
[77]. Saha, S., Saha, C., Haque, M. M., Alam, M. G. R., & Talukder, A. (2024). ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry. *IEEE Access*, *12*(NA), 4471-4484. https://doi.org/10.1109/access.2024.3349950
[78]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53
[79]. Sari, R. P., Febriyanto, F., & Adi, A. C. (2023). Analysis Implementation of the Ensemble Algorithm in Predicting Customer Churn in Telco Data: A Comparative Study. *Informatica*, *47*(7), NA-NA. https://doi.org/10.31449/inf.v47i7.4797
[80]. Sazzad, I., & Md Nazrul Islam, K. (2022). Project impact assessment frameworks in nonprofit development: a review of case studies from south asia. *American Journal of Scholarly Research and Innovation*, *1*(01), 270-294. https://doi.org/10.63125/eeja0t77
[81]. Šemrl, J., & Matei, A. (2017). BESC - Churn prediction model for effective gym customer retention. *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*, *NA*(NA), 1-3. https://doi.org/10.1109/besc.2017.8256379
[82]. Shaiful, M., Anisur, R., & Md, A. (2022). A systematic literature review on the role of digital health twins in preventive healthcare for personal and corporate wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373
[83]. Shoja, B. M., & Tabrizi, N. (2019). Customer Reviews Analysis With Deep Neural Networks for E-Commerce Recommender Systems. *IEEE Access*, *7*(NA), 119121-119130. https://doi.org/10.1109/access.2019.2937518
[84]. Sikri, A., Jameel, R., Idrees, S. M., & Kaur, H. (2024). Enhancing customer retention in telecom industry with machine learning driven churn prediction. *Scientific reports*, *14*(1), 13097-NA. https://doi.org/10.1038/s41598-024-63750-0
[85]. Singh, C., Dash, M. K., Sahu, R., & Kumar, A. (2023). Artificial intelligence in customer retention: a bibliometric analysis and future research framework. *Kybernetes*, *53*(11), 4863-4888. https://doi.org/10.1108/k-02-2023-0245
[86]. Sjarif, N. N. A., Azmi, N. F. M., Sarkan, H. M., Sam, S. M., & Osman, M. Z. (2020). Predicting Churn: How Multilayer Perceptron Method Can Help with Customer Retention in Telecom Industry. *IOP Conference Series: Materials Science and Engineering*, *864*(1), 012076-NA. https://doi.org/10.1088/1757-899x/864/1/012076
[87]. Subramanian, R. S., Yamini, B., Sudha, K., & Sivakumar, S. (2024). Ensemble-based deep learning techniques for customer churn prediction model. *Kybernetes*, *NA*(NA), NA-NA. https://doi.org/10.1108/k-08-2023-1516
[88]. Subrato, S. (2018). Resident's Awareness Towards Sustainable Tourism for Ecotourism Destination in Sundarban Forest, Bangladesh. *Pacific International Journal*, *1*(1), 32-45. https://doi.org/10.55014/pij.v1i1.38
[89]. Subrato, S., & Md, N. (2024). The role of perceived environmental responsibility in artificial intelligence-enabled risk management and sustainable decision-making. *American Journal of Advanced Technology and Engineering Solutions*, *4*(04), 33-56. https://doi.org/10.63125/7tjw3767
[90]. Swetha, P., & Dayananda, R. B. (2023). A customer churn prediction model in telecom industry using Improved_XGBoost. *International Journal of Cloud Computing*, *12*(2/3/4), 277-277. https://doi.org/10.1504/ijcc.2023.130903

[91]. Tahmina Akter, R. (2025). AI-driven marketing analytics for retail strategy: a systematic review of data-backed campaign optimization. *International Journal of Scientific Interdisciplinary Research*, 6(1), 28-59. https://doi.org/10.63125/0k4k5585

[92]. Tahmina Akter, R., & Abdur Razzak, C. (2022). The Role Of Artificial Intelligence In Vendor Performance Evaluation Within Digital Retail Supply Chains: A Review Of Strategic Decision-Making Models. *American Journal of Scholarly Research and Innovation*, *1*(01), 220-248. https://doi.org/10.63125/96jj3j86

[93]. Tao, J., Xiong, Y., Zhao, S., Wu, R., Shen, X., Lyu, T., Fan, C., Hu, Z., Zhao, S., & Pan, G. (2023). Explainable AI for Cheating Detection and Churn Prediction in Online Games. *IEEE Transactions on Games*, *15*(2), 242-251. https://doi.org/10.1109/tg.2022.3173399

[94]. Tonmoy, B., & Md Arifur, R. (2023). A Systematic Literature Review Of User-Centric Design In Digital Business Systems Enhancing Accessibility, Adoption, And Organizational Impact. *American Journal of Scholarly Research and Innovation*, *2*(02), 193-216. https://doi.org/10.63125/36w7fn47

[95]. Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*(NA), 60134-60149. https://doi.org/10.1109/access.2019.2914999

[96]. Umayaparvathi, V., & Iyakutti, K. (2012). Applications of Data Mining Techniques in Telecom Churn Prediction. *International Journal of Computer Applications*, *42*(20), 5-9. https://doi.org/10.5120/5814-8122

[97]. Utku, A., & Akcayol, M. A. (2020). Deep Learning Based Prediction Model for the Next Purchase. *Advances in Electrical and Computer Engineering*, *20*(2), 35-44. https://doi.org/10.4316/aece.2020.02005

[98]. Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, *38*(3), 2354-2364. https://doi.org/10.1016/j.eswa.2010.08.023

[99]. Vijaya, J., & Sivasankar, E. (2017). An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, *22*(5), 10757-10768. https://doi.org/10.1007/s10586-017-1172-1

[100]. Vijaya, J., & Sivasankar, E. (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. *Computing*, *100*(8), 839-860. https://doi.org/10.1007/s00607-018-0633-6

[101]. Vilone, G., & Longo, L. (2021). A Quantitative Evaluation of Global, Rule-Based Explanations of Post-Hoc, Model Agnostic Methods. *Frontiers in artificial intelligence*, *4*(NA), 717899-NA. https://doi.org/10.3389/frai.2021.717899

[102]. Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, *14*(NA), 100342-100342. https://doi.org/10.1016/j.rico.2023.100342

[103]. Wangunga, G., & Qu, Y. (2020). A Deep Learning based Customer Sentiment Analysis Model to Enhance Customer Retention and Loyalty in the Payment Industry. *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, *NA*(NA), 473-478. https://doi.org/10.1109/csci51800.2020.00086

[104]. Wu, F., Lyu, F., Ren, J., Yang, P., Qian, K., Gao, S., & Zhang, Y. (2024). Characterizing Internet Card User Portraits for Efficient Churn Prediction Model Design. *IEEE Transactions on Mobile Computing*, *23*(2), 1735-1752. https://doi.org/10.1109/tmc.2023.3241206

[105]. Xiahou, X., & Harada, Y. (2022). B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(2), 458-475. https://doi.org/10.3390/jtaer17020024

[106]. Zdravevski, E., Lameski, P., Apanowicz, C., & Ślęzak, D. (2020). From Big Data to business analytics: The case study of churn prediction. *Applied Soft Computing*, *90*(NA), 106164-NA. https://doi.org/10.1016/j.asoc.2020.106164

[107]. Zdziebko, T., Sulikowski, P., Sałabun, W., Przybyła-Kasperek, M., & Bąk, I. (2024). Optimizing Customer Retention in the Telecom Industry: A Fuzzy-Based Churn Modeling with Usage Data. *Electronics*, *13*(3), 469-469. https://doi.org/10.3390/electronics13030469

[108]. Zhang, X., Guo, F., Chen, T., Pan, L., Beliakov, G., & Wu, J. (2023). A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(4), 2188-2216. https://doi.org/10.3390/jtaer18040110

[109]. Zhu, B., Baesens, B., & Broucke, S. v. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, *408*(NA), 84-99. https://doi.org/10.1016/j.ins.2017.04.015

[110]. Zhu, B., Qian, C., vanden Broucke, S., Xiao, J., & Li, Y. (2023). A bagging-based selective ensemble model for churn prediction on imbalanced data. *Expert Systems with Applications*, *227*(NA), 120223-120223. https://doi.org/10.1016/j.eswa.2023.120223