

CUSTOMER CHURN PREDICTION USING ML TECHNIQUES: A REVIEW

**Ansari Mohd Armash¹, Chhipi Abdul Karim², Ansari Mohammad Sadeem³,
Zaibunnisa L. H. Malik⁴**

^{1,2,3}Computer Engineering M.H. Saboo Siddik Polytechnic Byculla, India

⁴Computer Engineering (HOD) M.H. Saboo Siddik Polytechnic Byculla, India.

ansariarmash007@gmail.com

abdulkarimchhipi@gmail.com

sadeem.rehan@gmail.com

zebamalik@yahoo.com

DOI: <https://www.doi.org/10.58257/IJPREMS36916>

ABSTRACT

This review examines the use of machine learning (ML) techniques in predicting customer churn across various industries, including telecommunications, financial services, and e-commerce. The study explores the effectiveness of different models, such as decision trees, support vector machines (SVM), and neural networks, in identifying at-risk customers based on behavioral, transactional, and demographic data. Ensemble models and hybrid approaches, such as those combining decision trees and logistic regression, demonstrate superior accuracy in predicting churn, particularly in handling large and complex datasets. However, challenges like model interpretability and ethical concerns around data privacy and bias remain significant barriers to widespread adoption. To address these issues, the review highlights recent advancements in explainable artificial intelligence (XAI) and profit-driven machine learning models, which aim to balance accuracy with transparency. The analysis concludes that while ML models offer substantial promise in improving customer retention strategies, further research is needed to enhance their applicability across diverse sectors and ensure they operate ethically and transparently.

1. INTRODUCTION

Customer churn prediction is a significant challenge for businesses across industries, as retaining customers is often more cost-effective than acquiring new ones. The ability to predict churn with high accuracy enables companies to take preemptive actions to retain at-risk customers, improving profitability and customer satisfaction. In recent years, machine learning (ML) has transformed churn prediction by enabling more accurate models that analyze complex customer behavior patterns and provide real-time insights. The primary goal of churn prediction models is to identify factors that lead to customer churn and help businesses devise strategies to retain these customers, particularly in high-churn sectors like telecommunications, finance, and e-commerce [3][12].

Machine learning techniques such as decision trees, support vector machines, and neural networks have shown effectiveness in processing large datasets and detecting patterns that traditional methods often miss. Manzoor et al. highlight that ML-based churn prediction models can outperform conventional statistical models, offering better precision and scalability in churn prediction [1]. Furthermore, the growing complexity of customer data—encompassing demographics, behavior, and transaction history—has driven the need for hybrid approaches that combine multiple ML techniques for improved accuracy and robustness [17][21].

In high-churn sectors like telecom and banking, predictive models have been successfully employed. Studies by Verbeke et al. and Coussement et al. demonstrate that data mining techniques applied to the telecommunications sector can provide deep insights into customer behavior patterns that contribute to churn [18][19]. Similarly, models incorporating neural networks, as studied by Phanindra et al., have been successfully applied to high-frequency datasets in banking and financial services, offering deeper insights into customer decision-making [13][5].

However, challenges persist in balancing model complexity with interpretability. Complex models such as deep learning, while highly accurate, are often criticized for being "black-box" models that make it difficult for businesses to interpret the reasons behind their predictions. Hybrid models and interpretable ML techniques, explored by Jiang et al., aim to provide both high accuracy and actionable insights, thus making these models more practical for real-world use [24][22]. The inclusion of explainable artificial intelligence (XAI) tools has also enhanced the transparency and applicability of these models, particularly in industries requiring regulatory compliance [25][9].

As the field evolves, future research is expected to focus on refining models for greater personalization and precision. The integration of advanced technologies, such as genetic algorithms [22] and cost-sensitive machine learning [23],

promises further improvements in customer churn prediction across industries. Addressing ethical concerns around data privacy and algorithmic bias, as noted by Gomase et al., will be crucial for ensuring the broader adoption of these predictive technologies [9].

2. METHODS

a) Key Terminologies

1. Customer Churn Prediction: Customer churn refers to the process of customers discontinuing their relationship with a company. Machine learning (ML) models are frequently employed to predict which customers are at risk of churning, using historical customer data such as transactions, interactions, and behavioral patterns [7][13].
2. Machine Learning (ML): ML techniques, including supervised learning algorithms such as decision trees, random forests, support vector machines (SVM), and neural networks, are commonly used in churn prediction. These models analyze large datasets and detect complex patterns that can help identify churn risk [18][26].
3. Ensemble Learning: This technique combines the predictive power of multiple models to improve accuracy. Methods such as bagging, boosting, and stacking are widely used in churn prediction to enhance model performance, particularly in dealing with unbalanced datasets [1][6].
4. Hybrid Models: Hybrid models integrate different machine learning techniques to capitalize on their strengths while mitigating individual weaknesses. For example, combining neural networks with decision trees has been shown to improve prediction accuracy in high-churn industries like telecommunications and e-commerce [5][17].

b) Search Strategy

To gather relevant literature, a systematic search was conducted across various academic databases, including IEEE Xplore, Springer, Elsevier, and ISJR. Keywords such as “customer churn prediction,” “machine learning,” “telecom churn,” and “banking churn” were used to retrieve papers published between 2010 and 2024. The search aimed to include studies that explored a range of industries, such as telecommunications [19][29], financial services [5][16], and e-commerce [15][27].

c) Selection Criteria

To ensure that the review captures recent advancements in customer churn prediction using machine learning, the following selection criteria were applied:

Relevance to Churn Prediction: Only studies that specifically focused on predicting customer churn using machine learning models were included. Papers examining broader applications of machine learning in customer retention were excluded [24][3].

Recency: To reflect the latest developments in ML technologies, studies published after 2010 were prioritized, with a specific focus on research from the last five years [2][22].

Diverse Industry Focus: Research spanning different industries, including telecommunications, banking, and e-commerce, was included to provide a holistic view of churn prediction methods [10][20].

Empirical and Theoretical Studies: Both theoretical frameworks and empirical studies applying ML models to real-world datasets were considered. This approach ensured that the review covered both model development and practical implementation [9][30].

d) Data Analysis

The selected papers were analyzed based on the machine learning models they employed, their performance metrics (such as accuracy, precision, and recall), and the types of datasets used (e.g., customer behavior data, financial transactions, or online interaction logs). Particular attention was given to the trade-off between model complexity and interpretability, a common issue in advanced machine learning models like deep learning [4][21]. Studies comparing the performance of different ML models, such as logistic regression, SVM, and random forests, were reviewed to highlight the most effective approaches for churn prediction in specific industries [11][14].

3. RESULTS

a) Performance of Machine Learning Models

Ensemble models, such as boosting and bagging techniques, have consistently demonstrated superior performance in churn prediction due to their ability to combine multiple weak models into a strong predictive model. For instance, the study by Verbeke et al. [18] applied boosting algorithms to telecommunications data, achieving high accuracy rates and offering robust predictions for high-churn customers. Similarly, Jiang et al. [24] reported that hybrid models combining decision trees with logistic regression produced high precision in predicting churn, particularly in the financial sector. Ensemble methods, especially stacking, showed a significant performance boost over single classifiers like logistic regression or random forests, as observed by Coussement et al. [19].

Neural networks, particularly deep learning models, also performed well, especially in handling large datasets and identifying complex patterns. Phanindra et al. [13] demonstrated that artificial neural networks (ANNs) could outperform traditional models when applied to high-frequency datasets in the banking sector. Similarly, the hybrid black-box classification model used by de Caingy et al. [21] offered improved accuracy for predicting churn while retaining interpretability—a major challenge in neural networks.

Support vector machines (SVMs) have shown success in handling high-dimensional data, particularly in the telecom industry. Yabas et al. [2] applied SVMs to telecom customer data, achieving high accuracy while maintaining model simplicity. However, the complexity of tuning SVM parameters and computational cost remains a limitation, as highlighted by Szeląg and Słowiński [20].

b) Application Across Industries

Telecommunications and banking sectors have been the most prominent industries for applying machine learning churn prediction models. Studies by Soni and Nelson [27] in the banking sector showed that machine learning techniques, particularly profit-driven models, helped financial institutions retain valuable customers by identifying those likely to churn based on transaction histories and financial behaviors. In contrast, Kesiraju and Deeplakshmi [3] demonstrated that dynamic customer behavior models could effectively predict churn in telecom, especially in regions with high customer turnover rates.

In the e-commerce sector, Shobana et al. [15] showed that machine learning-based business intelligence strategies could reduce customer churn by analyzing online customer behavior, such as shopping patterns and engagement metrics. The use of hybrid models in this sector was particularly effective in identifying at-risk customers and tailoring marketing efforts accordingly [15].

c) Model Interpretability and Real-World Challenges

Despite the advancements in accuracy and precision, the interpretability of machine learning models remains a challenge. Neural networks and ensemble models, while effective, often function as "black boxes," making it difficult for businesses to understand the reasoning behind predictions. To address this issue, models like the profit-driven logistic model developed by Stripling et al. [22] and the interpretable weighted classifier proposed by Jiang et al. [23] offer insights into the key features influencing churn, thereby improving trust in model predictions. Real-world implementation of these models also faces challenges related to data privacy and model generalizability. Lemos et al. [5] highlighted that while machine learning models performed well in predicting churn in financial institutions, issues such as data bias and privacy concerns could limit the broad application of these models. Moreover, Saleh and Saha [12] noted that models developed for one geographical or demographic context often underperform when transferred to different regions, underscoring the need for localized adaptations.

4. DISCUSSION

a) Model Interpretability Challenges

One of the most significant challenges in applying machine learning models for customer churn prediction is their interpretability. While models such as neural networks and ensemble methods (e.g., boosting, stacking) deliver high accuracy, they often act as "black-box" models, making it difficult for business stakeholders to understand why certain customers are predicted to churn. This lack of transparency limits the practical application of these models in decision-making processes, especially in sectors like finance and telecommunications, where understanding customer behavior is critical [13][21]. Hybrid approaches, such as those proposed by de Caingy et al. [21], attempt to address this by incorporating segmented interpretability, allowing businesses to gain insights into the factors driving churn without sacrificing accuracy.

b) Context-Specific Model Performance

The performance of machine learning models in churn prediction is highly dependent on the specific context in which they are applied. Models developed for one industry or region may underperform when transferred to another due to variations in customer behavior, demographics, and economic conditions. Saleh and Saha [12] found that machine learning models developed for Western markets failed to perform well in developing countries, where customers exhibited different churn triggers. This emphasizes the need for localized adaptations of models to ensure their effectiveness across diverse contexts [5]. Additionally, the banking sector has seen the successful use of profit-driven models that target high-value customers, as explored by Soni and Nelson [27], further underlining the importance of sector-specific model optimization.

c) Ethical and Privacy Concerns

Machine learning models rely heavily on sensitive customer data, raising significant ethical concerns related to data privacy and algorithmic bias. As Lemos et al. [5] point out, while machine learning models can provide valuable insights

into customer behavior, biases in the training data can lead to unfair predictions, particularly in industries like finance. Furthermore, privacy regulations such as the General Data Protection Regulation (GDPR) in the European Union impose strict requirements on how customer data is collected, stored, and used, limiting the widespread adoption of these models [9]. To address these concerns, Jiang et al. [23] propose the use of interpretable, bias-mitigating models that ensure fairness while delivering high accuracy.

d) The Role of Explainable AI (XAI) in Enhancing Trust

The integration of explainable artificial intelligence (XAI) techniques has emerged as a promising solution to improve the transparency and trustworthiness of machine learning models in churn prediction. XAI tools aim to make model decisions more interpretable, allowing businesses to understand why specific customers are predicted to churn. This is particularly important in sectors that require compliance with regulatory standards, such as finance and healthcare. Jiang et al. [23] highlight the effectiveness of XAI in balancing the trade-off between accuracy and interpretability, making models more actionable for business users. As XAI continues to evolve, it is expected to play a key role in increasing the adoption of machine learning models in churn prediction.

e) Future Directions for Churn Prediction Models

The future of churn prediction models lies in refining existing techniques to improve both accuracy and interpretability. Genetic algorithms and profit-driven models, such as those explored by Stripling et al. [22], offer significant potential for optimizing churn prediction by focusing on high-value customers. These techniques allow businesses to maximize profits by targeting retention efforts where they will have the most impact. Additionally, hybrid models that combine multiple machine learning techniques, such as decision trees and neural networks, are expected to continue evolving, offering both accuracy and the ability to handle complex datasets [24]. Moreover, future research should focus on mitigating bias and enhancing privacy protections to ensure that these models are both effective and ethically sound [5].

5. CONCLUSION

Machine learning has emerged as a crucial tool for predicting customer churn across industries, providing businesses with valuable insights that enable proactive customer retention strategies. Techniques such as ensemble models, neural networks, and support vector machines have shown superior performance in handling large, complex datasets and delivering accurate predictions. Studies across various sectors, including telecommunications [18][19], financial services [5][27], and e-commerce [15], demonstrate the versatility and effectiveness of these models in identifying at-risk customers. However, challenges remain, particularly around model interpretability and the ethical use of customer data. While deep learning and hybrid models achieve high accuracy, their "black-box" nature makes it difficult for businesses to understand the underlying drivers of churn, limiting their practical application [21][24]. Efforts to integrate explainable artificial intelligence (XAI) techniques and more interpretable models are crucial for improving trust and adoption in real-world settings [23]. Moreover, the transferability of models across different geographical and industry contexts remains limited, necessitating localized adaptations [12][5]. Ethical concerns, especially around data privacy and algorithmic bias, also present significant challenges. To ensure that machine learning models are fair and transparent, future research must focus on mitigating bias and enhancing data protection measures [9][22]. As businesses continue to rely on these models, advancements such as genetic algorithms and profit-driven techniques [22][23] offer exciting possibilities for refining churn prediction and improving customer retention strategies across industries. In summary, while machine learning models have proven highly effective in predicting customer churn, balancing accuracy, interpretability, and ethical considerations will be key to their future success and wider adoption. Moving forward, the development of more inclusive, interpretable, and adaptive machine learning models will be crucial to addressing the diverse needs of students globally. By focusing on ethical practices and transparency, the use of machine learning in education has the potential to revolutionize student retention and academic success across all levels of education [5][29][12].

6. REFERENCES

- [1] A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioner - Awais Manzoor, M. Atif Qureshi, Etain Kidney, Luca Longo, IEEE, 2024.
- [2] Customer Churn Prediction for Telecom Services Utku Yabas, Hakki Candan Cankaya, Turker Ince, IEEE, 2012.
- [3] Dynamic Churn Prediction using Machine Learning Algorithms - Predict your customer through customer behaviour - RajaGopal Kesiraju VLN, P. Deeplakshmi, IEEE, 2021.
- [4] A Review on Churn Prediction Modeling in Telecom Environment - Bhanuprakash P., IEEE, 2017.
- [5] Propensity to Customer Churn in a Financial Institution: A Machine Learning Approach - Renato Alexandre de Lima Lemos, Thiago Christiano Silva, Benjamin Miranda Tabak, Springer, 2022.

- [6] A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners - Awais Manzoor, M. Atif Qureshi, Etain Kidney, Luca Longo, IEEE, 2024.
- [7] A Comparative Model for Predicting Customer Churn using Supervised Machine Learning - Muchatibaya Adrin, David Fadaralika, ISJR, 2022.
- [8] A Study on Customer Churn Prediction - Suhel Malik, Siddhart Runwal, Yash Shah, Vishal Raut, Prof. S.S. Hire, IJRS, 2023.
- [9] Customer Churn and Appointment No-Show Analysis for Salon Management System - Hema Manoj Sudhir Gomase, Prof. Harshil T. Kanakia, IJRS, 2023.
- [10] Predicting and Factor Identification of Customer Churn in Business of B2C - Logachander B, N Gowri Vidhya, Likhith S, IJRS, 2023.
- [11] Customer Churn Prediction - Senthilnayaki B, Swetha M, Nivedha D, IJRS, 2021.
- [12] Customer Retention and Churn Prediction in the Telecommunication Industry: A Case Study on a Danish University - Sarkaft Saleh, Subrata Saha, IJRS, 2023.
- [13] Customer Churn Modelling using Artificial Neural Networks - T V S P Phanindra, G. Lakshmana Rao, Y Sai Ganesh, SV Ravi Shekhar, P Priyanka, IJSR, 2021.
- [14] Predictive Analysis by Studying Customer Behavior and Tools for Framing Marketing and Automation Strategies to Avoid Customer Churn - Mrs. Devibala Subramanian and Dr. R. Ravichandran, IJSR, 2020.
- [15] E-commerce customer churn prevention using machine learning-based business intelligence strategy - Shobana J, Ch. Gangadhar, Rakesh Kumar Arora, P.N. Renjith, J. Bamini, Yugendra devidas Chincholkar, Elsevier, 2023.
- [16] Propension to customer churn in a financial institution: a machine learning approach - Renato Alexandre de Lima Lemos, Thiago Christiano Silva, Benjamin Miranda Tabak, Springer, 2022.
- [17] Investigating customer churn in banking: a machine learning approach and visualization app for data science and management - Pahul Preet Singh, Fahim Islam Anik, Rahul Senapati, Arnav Sinha, Nazmus Sakib, Eklas Hossain, Elsevier, 2024.
- [18] New insights into churn prediction in the telecommunication sector: A profit-driven data mining approach - Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, Bart Baesens, Elsevier, 2012.
- [19] Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques - Kristof Coussement, Dirk Van den Poel, Elsevier, 2018.
- [20] Explaining and predicting customer churn by monotonic rules induced from ordinal data - Marcin Szeląg, Roman Słowiński, Elsevier, 2024.
- [21] Hybrid black-box classification for customer churn prediction with segmented interpretability analysis - Arno de Caingy, Koen W, DE Bock, Sam Verboven, Elsevier, 2024.
- [22] Profit maximizing logistic model for customer churn prediction using genetic algorithms - Eugen Stripling, Seppe VandenBroucke, Kotrein Antonio, Bart Baesens, Monique Snoeck, Elsevier, 2018.
- [23] Profit-driven weighted classifier with interpretable ability for customer churn prediction - Ping Jiang, Zhenkun Liu, Mohammad Zoynul Abedin, Jianzhou Wang, Wendong Yang, Qingli Dong, Elsevier, 2024.
- [24] Hybrid model for profit-driven churn prediction based on cost minimization and return maximization - Ping Jiang, Zhenkun Liu, Lifang Zhang, Jianzhou Wang, Elsevier, 2023.
- [25] Profit driven decision trees for churn prediction - Sebastiaan Höppner, Eugen Stripling, Bart Baesens, Seppe Vanden Broucke, Tim Verdonck, Elsevier, 2020.
- [26] A comparison of machine learning techniques for customer churn prediction - T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.Ch. Chatzisavvas, Elsevier, 2015.
- [27] Profit-Driven Churn Prediction using Machine Learning Techniques in Banking Sector - Pranshu Kumar Soni and Leema Nelson, ISJR, 2023.
- [28] Investigating customer churn in banking: a machine learning approach and visualization app for data science and management - Pahul Preet Singh, Fahim Islam Anik, Rahul Senapati, Arnav Sinha, Nazmus Sakib, Eklas Hossain, Elsevier, 2024.
- [29] Survey of CRM and Algorithmic Techniques used in Online Food Ordering System - Prof. Shailesh Galande, Mrunali Padwal, Karthi Dhanasekaran, Durga Late, Vivek Khatri, ISJR, 2023.
- [30] Profit Maximizing Logistic Model for Customer Churn Prediction Using Genetic Algorithms - Kristof Coussement, Dirk Van den Poel, IEEE, 2018.