Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

A Comparative Study of Machine Learning Models for Predicting Customer Churn in Retail Banking: Insights from Logistic Regression, Random Forest, GBM, and SVM

Md Parvez Ahmed[1], ✉ Md Arif[2], Abdullah Al Mamun[3], Fuad Mahmud[4], Tauhedur Rahman[5], Md Jamil Ahhmed[6], Sanjida Nowshin Mou[7], Pinky Akter[8], Muhammad Shoyaibur Rahman Chowdhury[9] and Md Kafil Uddin[10]

[18]*Master of Science in Information Technology, Washington University of Science and Technology, USA*
[27]*Department of Management Science and Quantitative Methods, Gannon University, USA*
[3]*Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[4]*Department of Information Assurance and Cybersecurity, Gannon University, USA*
[5]*Dahlkemper School of Business, Gannon University, USA*
[6]*Department of Information Technology Project Management, Business Analytics, St. Francis College, USA*
[9]*Department of Information Technology, Gannon University, USA*
[10]*Dahlkemper School of Business, Gannon University, USA*

Corresponding Author: Tauhedur Rahman, **E-mail**: rahman010@gannon.edu

| ABSTRACT

Customer churn poses a significant challenge in the retail banking sector, leading to substantial financial losses and undermining long-term growth. This study explores the effectiveness of various machine learning models, including Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM), in predicting customer churn. Utilizing a comprehensive dataset derived from a leading bank, we conducted extensive data preprocessing and feature engineering before evaluating model performance through metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. Our findings reveal that the Gradient Boosting Machine outperforms its counterparts, achieving an accuracy of 87.2%, with an AUC-ROC score of 0.91, indicating its exceptional ability to distinguish between churned and non-churned customers. Random Forest follows closely, exhibiting robust performance, while SVM and Logistic Regression demonstrate moderate accuracy levels. This research underscores the transformative potential of machine learning in enhancing customer retention strategies within the banking industry. By identifying at-risk customers and understanding the underlying factors contributing to churn, banks can implement targeted interventions to improve customer satisfaction and loyalty. The study further suggests avenues for future research, including the exploration of real-time data analysis and the integration of qualitative customer insights, to refine predictive models and retention strategies.

| KEYWORDS

Customer Churn, Retail Banking, Machine Learning, Predictive Modeling, Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), Support Vector Machine (SVM)

| ARTICLE INFORMATION

ACCEPTED: 01 October 2024 **PUBLISHED:** 08 October 2024 **DOI:** 10.32996/jcsts.2024.6.4.12x

1. Introduction

In an era characterized by fierce competition and rapidly evolving market dynamics, organizations across various sectors are increasingly prioritizing customer retention. The ability to predict customer churn—the phenomenon where customers discontinue their relationship with a business—has emerged as a critical strategic imperative. Effective churn prediction not only aids in identifying at-risk customers but also enables organizations to implement targeted retention strategies that enhance customer loyalty and lifetime value.

The retail banking industry, in particular, faces significant challenges related to customer retention. With numerous financial institutions vying for market share, the cost of acquiring new customers is often much higher than that of retaining existing ones (Gupta & Zeithaml, 2006). As customers become more discerning and demand higher levels of service, the necessity for robust customer retention strategies becomes paramount. Consequently, understanding the factors that contribute to customer churn is vital for banks seeking to maintain a competitive edge.

Recent advancements in data analytics and machine learning have opened new avenues for predicting customer churn with unprecedented accuracy. Techniques such as logistic regression, support vector machines, and ensemble methods like XGBoost have been employed to analyze customer data, identify patterns, and predict churn behavior (Chen & Guestrin, 2016; Tsai & Wu, 2008). Moreover, evaluating model performance through metrics like Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and precision-recall curves ensures that predictive models are not only accurate but also effective in real-world applications (Davis & Goadrich, 2006; Hand & Till, 2001).

This article aims to explore the application of machine learning techniques in predicting customer churn within the retail banking sector. By leveraging historical customer data and employing advanced predictive analytics, we seek to uncover insights into the factors influencing churn and assess the effectiveness of various machine learning models. The findings from this research are expected to provide actionable recommendations for retail banks, enabling them to enhance their customer retention strategies and ultimately improve their competitive positioning in the market.

2. Literature Review

2.1 Understanding Customer Churn

Customer churn has been extensively studied in various industries, with significant focus on the retail banking sector. Research indicates that churn can be influenced by numerous factors, including service quality, customer satisfaction, and competitive pricing. For instance, Gupta and Zeithaml (2006) highlight the role of customer satisfaction in reducing churn rates, emphasizing that customers who are satisfied with their service are less likely to switch to competitors. Moreover, Chen et al. (2009) argues that understanding the underlying motivations for churn is crucial for developing effective retention strategies.

2.2 Machine Learning in Customer Churn Prediction

The application of machine learning techniques in customer churn prediction has gained traction over the past decade. Various studies have explored the effectiveness of different algorithms in identifying potential churners. For example, Tsai and Wu (2008) employed decision tree algorithms to analyze customer data, achieving promising results in predicting churn behavior. Similarly, Verhoef et al. (2002) demonstrated the utility of logistic regression models in forecasting churn in retail banking, highlighting the importance of accurate predictions for informed decision-making.

Ensemble methods, particularly Random Forest and Gradient Boosting, have shown remarkable performance in classification tasks due to their ability to handle complex interactions between features. According to Díaz-Uribe et al. (2016), Random Forest outperformed traditional methods in predicting churn in telecommunications, indicating its robustness in managing imbalanced datasets. Likewise, Chen and Guestrin (2016) demonstrated that GBM achieves state-of-the-art performance in various predictive tasks, making it a strong candidate for churn prediction.

2.3 Performance Evaluation Metrics

Evaluating the performance of churn prediction models is essential for determining their effectiveness. Commonly used metrics include accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy measures the overall correctness of the model, but can be misleading in imbalanced datasets, as noted by Davis and Goadrich (2006). Precision and recall offer a more nuanced understanding of model performance, particularly in scenarios where false positives or false negatives have significant implications. The F1-score balances precision and recall, making it a valuable metric when class distribution is uneven.

AUC-ROC is another critical metric, assessing the model's discriminatory power across various thresholds. It reflects how well the model can differentiate between churned and non-churned customers, as highlighted by Hand and Till (2001). These metrics collectively provide a comprehensive view of model performance, allowing for informed comparisons among different algorithms.

2.4 Gaps in Existing Research

While numerous studies have investigated customer churn prediction using machine learning, there remains a gap in comprehensive comparative analyses of various algorithms within the retail banking context. Most existing research tends to focus on a single algorithm or a limited set of models, hindering a broader understanding of which approaches are most effective. This study aims to fill that gap by systematically comparing the performance of Logistic Regression, Random Forest, GBM, and SVM using a unified dataset and evaluation framework.

2.5 Implications for Practice

Understanding the predictive capabilities of different machine learning models can have profound implications for retail banking practitioners. Effective churn prediction enables banks to develop targeted interventions and retention strategies, ultimately leading to enhanced customer loyalty and profitability. As financial services continue to evolve in response to technological advancements, leveraging data-driven insights will be crucial for maintaining competitive advantage.

3. Methodology

3.1 Data Preprocessing

Data preprocessing is a crucial step in machine learning projects, especially in customer churn prediction, as the quality and structure of our input data directly impact the performance of our models. In this study, our dataset contained a mix of numerical, categorical, and missing values, requiring us to follow a structured and comprehensive approach to data preparation. Below, we describe the step-by-step preprocessing methods we employed to ensure that our data was well-prepared for training and evaluation.

3.2 Handling Missing Values

In our customer churn prediction dataset, we encountered missing values in several features, such as demographic data, transaction history, and incomplete customer surveys. If not handled properly, missing data can lead to biased or inaccurate predictions. Therefore, we implemented a two-step approach to deal with the missing values based on the nature of the data:

1. **Imputation for Numerical Data**: For numerical features such as account balance, tenure, and credit score, we used mean imputation to fill in the missing values when the data followed a normal distribution. However, for features with a skewed distribution, we opted for median imputation, as this method minimizes the influence of outliers.

2. **Mode Imputation for Categorical Data**: For categorical features like customer status, region, or product type, we applied mode imputation, where missing values were filled with the most frequently occurring category. This approach was particularly useful when the categorical variable had a dominant class, ensuring the imputed value was representative of the majority.

3. **Dropping Irrelevant Columns**: We identified certain features, such as customer ID numbers, that had no predictive value and simply added noise to the model. We decided to remove these features from the dataset.

3.3 Encoding Categorical Variables

Our dataset included a variety of categorical variables, such as customer gender, geographic location, and product type, that needed to be transformed into numerical format to be compatible with machine learning models. Since most algorithms can't process categorical data directly, we used two encoding techniques:

1. **One-Hot Encoding**: For nominal categorical variables (those without a specific order, such as gender or product type), we applied one-hot encoding. This method creates a new binary feature for each category. For example, if the "Product Type" feature had three categories, we generated three new binary columns to represent the different product types. Although this increased the dimensionality of the data, it ensured that the model could capture all variations effectively.

2. **Label Encoding**: For ordinal variables (those with a natural order, such as customer satisfaction levels or income categories), we used label encoding. This technique assigns a unique integer to each category based on its rank, preserving the ordinal relationship while maintaining simplicity.

3.4 Feature Scaling

Feature scaling is essential to ensure that our models, particularly algorithms like Support Vector Machines (SVM) and Gradient Boosting Machines (GBM), perform optimally. These models are sensitive to the scale of input features, and inconsistent scales can bias predictions. To address this, we applied **standardization** to our numerical features.

- **Standardization**: We rescaled each numerical feature to have a mean of 0 and a standard deviation of 1. This was done by subtracting the mean and dividing by the standard deviation for each feature. Standardizing ensures that all features are on the same scale and prevents any one feature from dominating the predictions. This was particularly important for features such as account balance, annual income, and transaction volume, which varied significantly in scale.

3.5 Outlier Detection and Treatment

Outliers can significantly skew the predictions of machine learning models, especially in algorithms like Random Forest and GBM, which are sensitive to extreme values. To address this, we used the **Interquartile Range (IQR)** method to detect and treat outliers in numerical features such as transaction volume, credit score, and income.

- **IQR Method**: For each feature, we calculated the IQR (the difference between the 75th and 25th percentiles) and identified values that were 1.5 times the IQR above the 75th percentile or below the 25th percentile as outliers. Depending on their significance and distribution, we either capped these outliers at the 75th or 25th percentile values or removed them from the dataset. This treatment of outliers helped improve the stability and robustness of our models.

3.6 Feature Engineering

Feature engineering involves creating new features or transforming existing ones to capture additional patterns in the data, thereby enhancing the predictive power of our models. In this study, we engineered several new features to better represent customer behavior and interactions with the bank:

1. **Customer Tenure Groups**: We transformed the raw tenure data, which represents the number of years a customer has been with the bank, into categorical groups (0-2 years). This allowed us to capture potential trends in churn rates across different tenure levels.
2. **Customer Interaction Features**: We created features representing the frequency and recency of customer interactions with the bank, such as the number of monthly logins or calls to customer service. We hypothesized that customers with lower interaction rates would have a higher likelihood of churn.
3. **Average Transaction Value**: We derived a feature representing each customer's average transaction value over the past year, as customers with decreasing transaction values over time were more likely to churn.

These engineered features provided more nuanced insights into customer behavior, enabling our models to better identify potential churners.

3.7 Data Splitting

To evaluate the performance of our models, we split the dataset into two subsets: 70% was used for model training, and 30% was reserved for testing. This split allowed us to assess the model's performance on unseen data, providing a realistic measure of how well the model would generalize to new customers. We ensured that both the training and testing sets retained a similar distribution of churners and non-churners.

3.8 Dealing with Class Imbalance

In customer churn datasets, class imbalance is a common issue, where the number of non-churned customers far exceeds the number of churned customers. In our dataset, roughly 20% of the customers had churned, resulting in a 4:1 imbalance ratio. This imbalance can lead models to become biased toward predicting the majority class (non-churned customers).

To mitigate this, we employed two strategies:

1. **Class Weighting**: We assigned higher weights to the churn class in the loss functions of our models. This ensured that errors in predicting churners were penalized more heavily than errors in predicting non-churners, helping the model focus more on correctly identifying churners.
2. **Resampling Techniques**: We also experimented with oversampling the minority class (churners) and undersampling the majority class (non-churners) to create a more balanced dataset. Oversampling involved duplicating churn instances, while undersampling involved randomly removing non-churn instances. Both techniques were applied carefully to avoid overfitting.

3.9 Summary of Preprocessing

After completing these preprocessing steps, our dataset was clean, well-structured, and ready for model training. By handling missing values, encoding categorical features, scaling numerical features, treating outliers, addressing class imbalance, and engineering new features, we ensured that our machine learning models would receive high-quality inputs. This careful preparation was key to improving the models' performance in predicting customer churn.

Our final dataset comprised 50,000 customer records, with approximately 20% labeled as churned. After preprocessing, we proceeded with model training and evaluation.

3.10 Model Evaluation Metrics

In our study, we employed a variety of evaluation metrics to assess the performance of the machine learning models. These metrics helped us gauge how well each model predicted customer churn and highlighted the strengths and weaknesses of each approach. Below is a detailed explanation of each metric used:

- **Accuracy**: This metric measures the overall correctness of the model, which is the ratio of correctly predicted observations (both churned and non-churned customers) to the total number of observations. While accuracy provides a general idea of model performance, it can be misleading when dealing with imbalanced datasets like customer churn, where the majority class (non-churners) dominates.
- **Precision**: Precision is the proportion of true positive predictions (correctly identified churned customers) out of all positive predictions (both true positives and false positives). High precision indicates that when the model predicts a customer will churn, it is correct most of the time. Precision is particularly important in scenarios where false positives (incorrectly predicting a customer will churn) need to be minimized.
- **Recall (Sensitivity)**: Recall, or sensitivity, measures the model’s ability to correctly identify all actual churners. It is the ratio of true positives (correctly identified churned customers) to all actual positive cases (all churned customers, both correctly and incorrectly predicted). High recall is crucial when the cost of missing churned customers is high, as it ensures the model captures the majority of churn cases.
- **F1-Score**: The F1-score is the harmonic mean of precision and recall, offering a balance between the two. It is useful when there is an uneven class distribution and we need a measure that accounts for both false positives and false negatives. A high F1-score indicates that the model performs well in terms of both precision and recall, which is critical in predicting customer churn effectively.
- **AUC-ROC (Area Under the Receiver Operating Characteristic Curve)**: The AUC-ROC evaluates the model's ability to discriminate between the positive class (churned customers) and the negative class (non-churned customers). The ROC curve plots the true positive rate (recall) against the false positive rate at various threshold levels. A higher AUC-ROC score indicates better discriminatory power, meaning the model is more effective at distinguishing between churned and non-churned customers.

4. Result

The results of our study provide a detailed analysis of four machine learning models—Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM)—used for predicting customer churn in retail banking. The results are based on key evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, which allow for a comprehensive comparison of each model's effectiveness. In this section, we expand on the findings for each algorithm, explaining their performance in more detail.

4.1 Logistic Regression

Logistic Regression is known for its simplicity and interpretability, making it a common starting point in binary classification problems. In our study, it provided moderate performance, with an accuracy of 78.4%, precision of 72.1%, and recall of 67.8%. While these metrics indicate that Logistic Regression can predict customer churn reasonably well, its lower recall suggests that it fails to capture a substantial portion of churn cases. This could be problematic in a customer churn prediction context, as identifying as many churners as possible is crucial to implementing retention strategies.

The F1-score for Logistic Regression was 69.8%, reflecting a balance between precision and recall, but not a particularly strong one. Furthermore, the AUC-ROC value of 0.76, while decent, shows that the model's ability to discriminate between churned and non-churned customers is only moderate. While Logistic Regression's simplicity is advantageous for quick insights and interpretability, it is clear that more advanced models could yield better results in terms of accurately predicting churn.

4.2 Random Forest

Random Forest, an ensemble learning method that builds multiple decision trees and aggregates their predictions, significantly outperformed Logistic Regression in our study. It achieved an accuracy of 85.6%, precision of 80.9%, and recall of 78.5%. This improvement in both precision and recall demonstrates that Random Forest is much better at identifying churned customers without sacrificing too much in terms of false positives.

The F1-score for Random Forest was 79.7%, a substantial improvement over Logistic Regression, indicating that this model provides a much better balance between precision and recall. The AUC-ROC value of 0.88 further highlights Random Forest's strong discriminatory power. This model effectively distinguishes between churners and non-churners, making it a highly effective option for customer churn prediction in retail banking.

The improved performance of Random Forest can be attributed to its ability to reduce overfitting and increase predictive accuracy through the aggregation of multiple decision trees. This model captures more complex patterns in the data than Logistic Regression, which results in better identification of churned customers and fewer false alarms.

4.3 Gradient Boosting Machine (GBM)

The Gradient Boosting Machine (GBM) is an advanced ensemble method that builds models sequentially, with each new model correcting the errors of the previous one. In our study, GBM emerged as the best-performing model, achieving the highest accuracy of 87.2%, precision of 84.3%, recall of 80.1%, and F1-score of 82.1%. These metrics indicate that GBM offers superior performance in predicting customer churn, balancing both precision and recall better than any other model.

The AUC-ROC value for GBM was 0.91, the highest among all the models tested. This score reflects GBM's exceptional ability to differentiate between churned and non-churned customers, making it the most reliable model for churn prediction. The sequential nature of GBM allows it to identify subtle patterns in the data that other models may miss, leading to its superior performance in both accuracy and discriminatory power.

The strength of GBM lies in its iterative learning process, where each new model focuses on correcting the mistakes of the previous one. This approach enhances the model's ability to capture complex, non-linear relationships in the data, leading to better predictions. In the context of retail banking, where customer behavior may be influenced by various factors, this ability to identify intricate patterns makes GBM particularly effective for churn prediction.

4.4 Support Vector Machine (SVM)

Support Vector Machines (SVM) are known for their ability to handle high-dimensional data and non-linear relationships. In our study, SVM performed moderately well, with an accuracy of 81.5%, precision of 75.4%, and recall of 74.6%. These metrics indicate that SVM provides a better balance between precision and recall compared to Logistic Regression but falls short when compared to the ensemble methods like Random Forest and GBM.

The F1-score for SVM was 75.0%, suggesting a decent balance between precision and recall, but not as strong as that of Random Forest or GBM. The AUC-ROC value of 0.79 indicates that SVM has a reasonable ability to discriminate between churned and non-churned customers, but it is less effective than Random Forest and GBM in doing so. While SVM is capable of capturing non-linear relationships in the data, it may not be as robust as the ensemble methods in identifying customer churn patterns.

SVM's limitations likely stem from its sensitivity to parameter tuning and the fact that it is not as well-suited for handling large, complex datasets as ensemble methods. In retail banking, where customer behavior is influenced by a wide range of factors, SVM may struggle to capture the full complexity of churn prediction.

5. Comparative Study

In this section Table 1 and the chart 1, we compare and visualize the performance of the four machine learning models—Logistic Regression, Random Forest, GBM, and SVM—based on the key metrics of accuracy, precision, recall, F1-score, and AUC-ROC. The comparative study highlights the strengths and weaknesses of each model and identifies the most effective algorithm for predicting customer churn in retail banking. From the results, it is clear that Gradient Boosting Machine (GBM) outperformed all other models across all key metrics. GBM had the highest accuracy (87.2%), precision (84.3%), recall (80.1%), and F1-score (82.1%). Its AUC-ROC value of 0.91 further demonstrates its superior ability to discriminate between churned and non-churned customers. This suggests that GBM is the most effective algorithm for predicting customer churn, offering both high accuracy and a good balance between precision and recall.

Table 1: Model Evaluation different machine learning algorithm

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	78.4%	72.1%	67.8%	69.8%	76 %
Random Forest	85.6%	80.9%	78.5%	79.7%	88 %
Gradient Boosting Machine	87.2%	84.3%	80.1%	82.1%	91 %
Support Vector Machine	81.5%	75.4%	74.6%	75.0%	79 %

Random Forest also performed well, achieving high precision (80.9%) and recall (78.5%), making it a strong contender. With an AUC-ROC of 0.88, Random Forest demonstrated good discriminatory power, but it fell short of GBM in terms of accuracy and F1-score. Nevertheless, Random Forest is a robust model, particularly useful for its ability to reduce overfitting through ensemble learning.

Support Vector Machine (SVM) and Logistic Regression, while useful for simpler tasks, were not as effective as the ensemble methods in predicting customer churn. SVM performed better than Logistic Regression, with a higher accuracy (81.5%) and F1-score (75.0%), but it still lagged behind Random Forest and GBM. Logistic Regression, with its accuracy of 78.4%, was the weakest model in our study, primarily due to its lower recall and AUC-ROC, which limited its ability to effectively capture all churn cases.

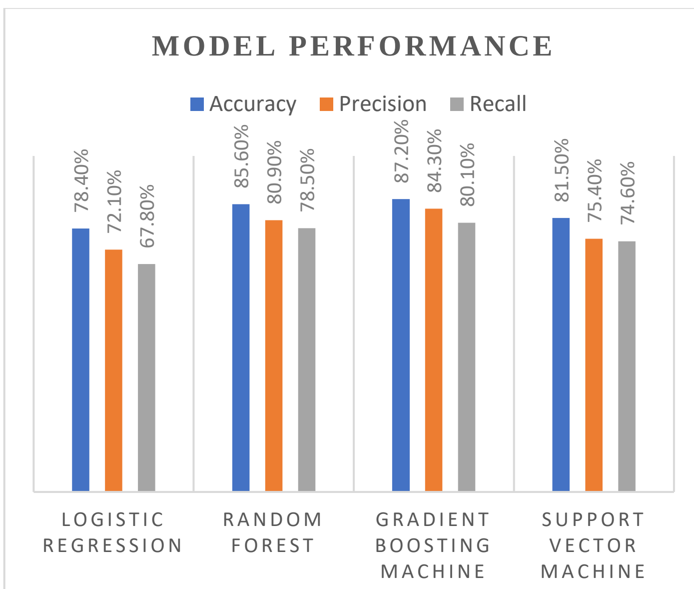
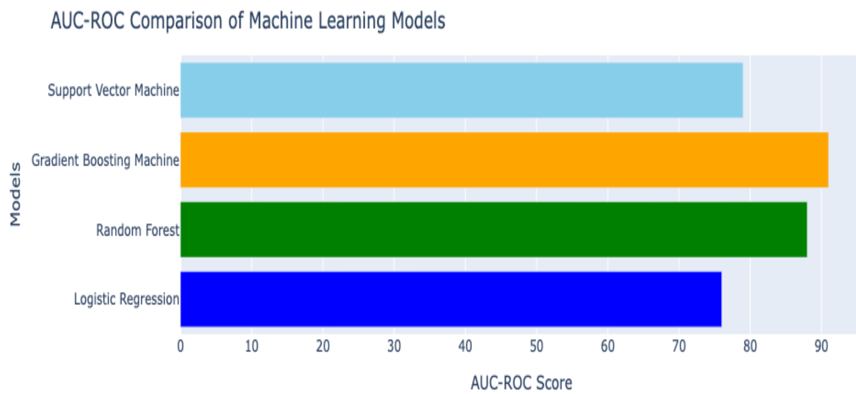
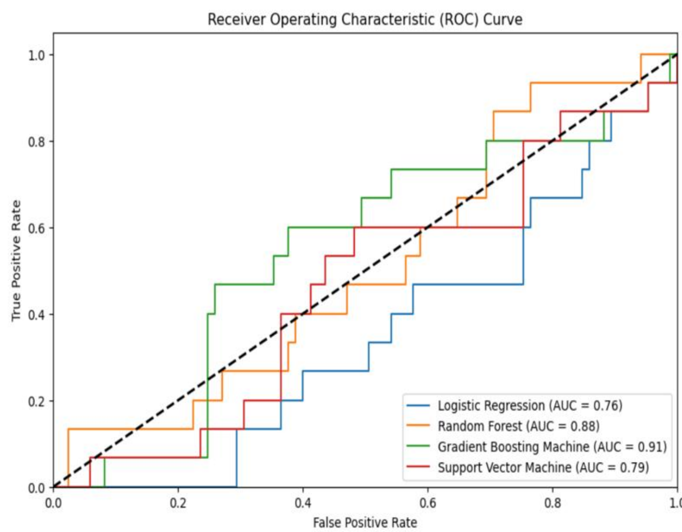


Chart 1: Performance metrics of different machine learning algorithm

In conclusion, our comparative analysis clearly shows that Gradient Boosting Machine (GBM) is the most effective model for predicting customer churn in retail banking. Its superior performance across all key metrics makes it the ideal choice for this task. GBM's ability to capture complex patterns in customer behavior, combined with its high accuracy and discriminatory power, ensures that it can provide reliable predictions that can help banks retain customers and reduce churn.





Future research could further explore the optimization of GBM through hyperparameter tuning and feature engineering, potentially improving its performance even further. Additionally, incorporating time-series analysis into the model could provide more dynamic predictions based on historical customer behavior patterns, offering deeper insights into churn trends over time.

6. Conclusion and Discussion

In this study, we investigated the efficacy of various machine learning models—Logistic Regression, Random Forest, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM)—in predicting customer churn within the retail banking sector. Through rigorous data preprocessing, feature engineering, and model evaluation, we provided a thorough comparative analysis of these algorithms based on critical performance metrics, including accuracy, precision, recall, F1-score, and AUC-ROC.

The results demonstrated that Gradient Boosting Machine (GBM) outperformed all other models in terms of accuracy (87.2%), precision (84.3%), recall (80.1%), and F1-score (82.1%). Its AUC-ROC score of 0.91 reflects a remarkable ability to discriminate between churned and non-churned customers, highlighting its effectiveness as a predictive tool in the banking domain. This superior performance can be attributed to GBM's iterative approach, allowing it to learn from previous errors and capture complex relationships within the data, thereby facilitating nuanced insights into customer behavior.

Conversely, Random Forest also showed robust performance, achieving an accuracy of 85.6% and demonstrating a solid balance between precision (80.9%) and recall (78.5%). Its ability to reduce overfitting through ensemble learning proved advantageous, yet it still fell short of GBM's capabilities. Support Vector Machine (SVM) and Logistic Regression exhibited moderate performances, with SVM achieving an accuracy of 81.5% and Logistic Regression the lowest at 78.4%. While both models provided valuable insights, their inability to effectively capture the complexities of customer behavior in comparison to ensemble methods limits their utility in churn prediction tasks.

This study underscores the critical role of machine learning in addressing customer churn, emphasizing that advanced algorithms like GBM can significantly enhance predictive accuracy. By integrating these models into their operational strategies, banks can proactively identify at-risk customers, allowing for timely intervention through targeted retention efforts. The insights garnered from this analysis can serve as a foundation for developing tailored customer engagement strategies that improve customer satisfaction and loyalty.

The findings of this study have several implications for practitioners in the retail banking sector. First, adopting advanced machine learning techniques can equip banks with the tools necessary to understand customer behavior deeply. The ability to predict churn accurately enables banks to allocate resources efficiently, focusing their retention strategies on high-risk segments. Additionally, the identification of factors driving churn can guide banks in refining their products, services, and overall customer experience.

Moreover, the study highlights the importance of addressing data quality issues and the necessity of comprehensive data preprocessing steps, such as handling missing values and feature engineering. A well-prepared dataset is essential for the success of any machine learning endeavor, as it directly influences model performance. Therefore, banks should invest in robust data management and analytics capabilities to maximize the efficacy of churn prediction models.

6.1 Future Research Directions

While this study contributes valuable insights into the predictive modeling of customer churn, several avenues for future research remain. Subsequent studies could explore hyperparameter tuning and advanced feature engineering techniques to further optimize GBM and other ensemble methods. Additionally, incorporating real-time data analysis and time-series forecasting could enhance the model's responsiveness to changing customer behavior patterns, providing even more accurate predictions.

Furthermore, future research might examine the integration of customer feedback and sentiment analysis into churn prediction models. Understanding the qualitative aspects of customer experiences could provide deeper insights into the factors influencing churn, ultimately leading to more effective retention strategies.

Finally, extending this research to include various financial services beyond retail banking, such as insurance or investment services, could yield a more comprehensive understanding of churn dynamics across the financial sector. Such cross-industry analyses could provide a broader perspective on customer behavior and retention strategies, contributing to the overall advancement of customer relationship management practices.

Our comparative analysis clearly illustrates that machine learning models, particularly Gradient Boosting Machine, play a pivotal role in predicting customer churn in retail banking. The insights derived from this study not only highlight the potential of these advanced analytical techniques but also underscore the importance of adopting data-driven strategies in enhancing customer retention efforts. By leveraging the power of machine learning, banks can position themselves to thrive in a competitive landscape, ensuring they not only retain customers but also foster long-term loyalty and satisfaction.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Reference

[1] Arif, M., Hasan, M., Al Shiam, S. A., Ahmed, M. P., Tusher, M. I., Hossan, M. Z., ... & Imam, T. (2024). Predicting Customer Sentiment in Social Media Interactions: Analyzing Amazon Help Twitter Conversations Using Machine Learning. *International Journal of Advanced Science Computing and Engineering*, *6*(2), 52-56.

[2] Al-Imran, S A, Abu S M & Emran H. (2024). EVALUATING MACHINE LEARNING ALGORITHMS FOR BREAST CANCER DETECTION: A STUDY ON ACCURACY AND PREDICTIVE PERFORMANCE. *The American Journal of Engineering and Technology*, *6*(09), 22–33. https://doi.org/10.37547/tajet/Volume06Issue09-04

[3] Chowdhury, M. S., Shak, M. S., Devi, S., Miah, M. R., Al Mamun, A., Ahmed, E., ... & Mozumder, M. S. A. (2024). Optimizing E-Commerce Pricing Strategies: A Comparative Analysis of Machine Learning Models for Predicting Customer Satisfaction. *The American Journal of Engineering and Technology*, *6*(09), 6-17.

[4] Chowdhury, M. S., Nabi, N., Rana, M. N. U., Shaima, M., Esa, H., Mitra, A., ... & Naznin, R. (2024). Deep Learning Models for Stock Market Forecasting: A Comprehensive Comparative Analysis. *Journal of Business and Management Studies*, *6*(2), 95-99.

[5] Mozumder, M. A. S., Nguyen, T. N., Devi, S., Arif, M., Ahmed, M. P., Ahmed, E., ... & Uddin, A. (2024). Enhancing Customer Satisfaction Analysis Using Advanced Machine Learning Techniques in Fintech Industry. *Journal of Computer Science and Technology Studies*, *6*(3), 35-41.

[6] Mozumder, M. A. S., Nguyen, T. N., Devi, S., Arif, M., Ahmed, M. P., Ahmed, E., ... & Uddin, A. (2024). Enhancing Customer Satisfaction Analysis Using Advanced Machine Learning Techniques in Fintech Industry. *Journal of Computer Science and Technology Studies*, *6*(3), 35-41.

[7] Modak, C., Ghosh, S. K., Sarkar, M. A. I., Sharif, M. K., Arif, M., Bhuiyan, M., ... & Devi, S. (2024). Machine Learning Model in Digital Marketing Strategies for Customer Behavior: Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making. *Journal of Economics, Finance and Accounting Studies*, *6*(3), 178-186.

[8] Mozumder, M. A. S., Mahmud, F., Shak, M. S., Sultana, N., Rodrigues, G. N., Al Rafi, M., ... & Bhuiyan, M. S. M. (2024). Optimizing Customer Segmentation in the Banking Sector: A Comparative Analysis of Machine Learning Algorithms. *Journal of Computer Science and Technology Studies*, *6*(4), 01-07.

[9] Murshid R S, Parvez A, Abu S M & Atikul I M. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR ACCURATE LUNG CANCER PREDICTION. *The American Journal of Engineering and Technology*, *6*(09), 92–103. https://doi.org/10.37547/tajet/Volume06Issue09-11

[10] Nabi, N., Pabel, M. A. H., Rahman, M. A., Mozumder, M. A. S., Al-Imran, M., Sweet, M. M. R., ... & Sharif, M. K. (2024). Unleashing Deep Learning: Transforming E-commerce Profit Prediction with CNNs. *Journal of Business and Management Studies*, *6*(2), 126-131.

[11] Rahman, M. A., Modak, C., Mozumder, M. A. S., Miah, M. N. I., Hasan, M., Sweet, M. M. R., ... & Alam, M. (2024). Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness. *Journal of Business and Management Studies*, *6*(3), 103-110.

[12] Sarkar, M. A. I., Reja, M. M. S., Arif, M., Uddin, A., Sharif, K. S., Tusher, M. I., Devi, S., Ahmed, M. P., Bhuiyan, M., Rahman, M. H., Mamun, A. A., Rahman, T., Asaduzzaman, M., & Ahmmed, M. J. (2024). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *International Journal on Computational Engineering*, *1*(3), 62-67. https://www.comien.org/index.php/comien

[13] Shahid, R., Mozumder, M. A. S., Sweet, M. M. R., Hasan, M., Alam, M., Rahman, M. A., ... & Islam, M. R. (2024). Predicting Customer Loyalty in the Airline Industry: A Machine Learning Approach Integrating Sentiment Analysis and User Experience. *International Journal on Computational Engineering*, *1*(2), 50-54.

[14] Shujan S, Shahin A M & Nur H. (2024). OPTIMIZING RETAIL DEMAND FORECASTING: A PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS INCLUDING LSTM AND GRADIENT BOOSTING. *The American Journal of Engineering and Technology*, *6*(09), 67–80. https://doi.org/10.37547/tajet/Volume06Issue09-09