The American Journal of
Engineering and Technology

ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 22-33
DOI 10.37547/tajet/Volume07Issue04-04

Check for updates

Comparative Analysis of Machine Learning Models for Credit Risk Prediction in Banking Systems

Safayet Hossain

Master of Science in Cybersecurity, Washington University of Science and Technology, USA

Ashadujjaman Sajal

Department of Management Science and Quantitative Methods, Gannon University, USA

Sakib Salam Jamee

Department of Management Information Systems, University of Pittsburgh, PA, USA

Sanjida Akter Tisha

Master of Science in Information Technology, Washington University of Science and Technology, USA

Md Tarake Siddique

Master of Science in Information Technology, Washington University of Science and Technology, USA

Md Omar Obaid

Department of Business Analytics, California State Polytechnic University Pomona, CA, USA

MD Sajedul Karim Chy

Department of Business Administration, Washington University of Science and Technology, USA

Md Sayem Ul Haque

MBA in Business Analytics, Gannon University, USA

OPEN ACCESS

SUBMITED 23 February 2025
ACCEPTED 25 March 2025
PUBLISHED 08 April 2025
VOLUME Vol.07 Issue04 2025

CITATION

Safayet Hossain, Ashadujjaman Sajal, Sakib Salam Jamee, Sanjida Akter Tisha, Md Tarake Siddique, Md Omar Obaid, MD Sajedul Karim Chy, & Md Sayem Ul Haque. (2025). Comparative Analysis of Machine Learning Models for Credit Risk Prediction in Banking Systems. The American Journal of Engineering and Technology, 7(04), 22–33. https://doi.org/10.37547/tajet/Volume07Issue04-04

Abstract: The increasing complexity of credit risk management in banking systems has led to the adoption of machine learning techniques to improve the prediction of loan defaults. This study evaluates and compares the performance of several machine learning models—Logistic Regression, Random Forest, Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Neural Networks—in predicting credit risk. The models were tested on a comprehensive dataset containing demographic, financial, and historical loan data. Performance was assessed based on accuracy, precision, recall, F1-score, AUC, and confusion matrix analysis. The results indicate that Gradient Boosting (XGBoost) outperformed the other models with the highest accuracy (88.7%), precision (89.5%), recall (80.3%), and AUC (91.3%), demonstrating its superior ability to predict loan defaults and manage credit risk effectively. Random Forest followed closely in performance, while Logistic Regression showed solid results with a focus on interpretability. Neural Networks and SVM performed well in accuracy but were more resource-intensive and less interpretable. The study concludes that Gradient Boosting (XGBoost) is the most suitable model for large-scale credit risk management due to its balance of high predictive power and ability to handle complex, imbalanced datasets. However, the choice of model should consider computational resources, interpretability requirements, and specific operational constraints of the banking institution.

Keywords: Machine learning, credit risk management, loan default prediction, Gradient Boosting, XGBoost, Random Forest, Logistic Regression, Support Vector Machines, Neural Networks, model comparison, predictive accuracy, banking systems.

Introduction: Credit risk management plays a crucial role in the stability and profitability of financial institutions. With the increasing volume of financial transactions and the complexity of borrower profiles, it has become essential for banks to develop accurate and efficient systems for predicting the likelihood of loan defaults. Traditional methods of credit risk assessment, such as statistical models and manual underwriting, have proven to be less effective in handling large-scale data and complex patterns that emerge from customer behaviors. As a result, financial institutions have increasingly turned to machine learning (ML) models, which offer the ability to process vast amounts of data and uncover intricate relationships between variables that may not be immediately obvious.

Machine learning algorithms, including Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Neural Networks, have demonstrated their potential to improve credit risk prediction by providing more accurate and reliable insights compared to traditional methods. These models can analyze diverse datasets, ranging from demographic information and financial histories to behavioral patterns, and generate predictions that aid decision-making in credit approval processes. However, the challenge remains in selecting the most suitable model for real-world applications, particularly when considering factors such as interpretability, computational efficiency, and performance in the context of banking systems.

This study aims to evaluate and compare the performance of various machine learning models in predicting credit risk, with a focus on their real-world applicability in banking systems. The models assessed include Logistic Regression, Random Forest, XGBoost, SVM, and Neural Networks, and their performance will be evaluated based on key metrics such as accuracy, precision, recall, F1-score, and AUC.

Literature Review

The application of machine learning in credit risk management has been a topic of growing interest in recent years, driven by the increasing availability of large datasets and the need for more accurate predictive models. Researchers have explored various machine learning techniques to improve the efficiency and accuracy of credit risk prediction.

One of the most widely used methods in credit risk modeling is Logistic Regression, which has been a cornerstone of statistical modeling in financial risk management for decades. Logistic Regression provides a simple yet interpretable model for binary classification problems, such as predicting whether a borrower will default on a loan. However, its performance can be limited when dealing with non-linear relationships and complex data (Chorafas, 2017). Despite these limitations, Logistic Regression remains a popular choice for simpler datasets due to its interpretability and ease of implementation.

In contrast, tree-based algorithms such as Random Forest and Gradient Boosting have gained significant traction in recent years due to their ability to handle large datasets and complex relationships. Random Forest, an ensemble learning method, builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. Several studies have demonstrated its effectiveness in credit risk modeling, with higher accuracy and better handling of missing data compared to traditional methods (Breiman, 2001). Similarly, Gradient Boosting, particularly the XGBoost implementation, has become one of the most popular algorithms for credit scoring. Its boosting mechanism, which sequentially builds trees to correct the errors of previous models, has been shown to outperform other algorithms in terms of accuracy and predictive power (Chen & Guestrin, 2016).

Support Vector Machines (SVM) have also been applied to credit risk management, particularly for their ability to handle high-dimensional data and non-linear decision boundaries. SVM has been found to perform well in identifying complex patterns in credit data, especially when combined with kernel methods (Cortes & Vapnik, 1995). However, SVMs require careful tuning of hyperparameters and can be computationally expensive, which may limit their scalability in large-scale banking systems.

Neural Networks, particularly deep learning models, have emerged as a promising technique for credit risk prediction due to their ability to capture intricate patterns in large and complex datasets. Several studies have demonstrated the superior performance of neural networks compared to traditional machine learning models in predicting loan defaults (Yao & Jiang, 2019). However, deep learning models often suffer from a lack of interpretability, which may be a concern for regulatory compliance in banking applications. Furthermore, neural networks require substantial computational resources and training time, making them less practical for smaller institutions.

Despite the growing adoption of machine learning techniques, challenges remain in integrating these models into real-world banking systems. The choice of model depends on several factors, including the size and complexity of the dataset, the interpretability of the model, and the computational resources available. Therefore, it is essential to compare the performance of different machine learning models to determine which one is most suitable for credit risk management in real-world banking applications.

METHODOLOGY

Dataset Collection

The foundation of any predictive model lies in the dataset that is used to train and validate it. For the credit risk management problem, a high-quality dataset containing historical records of borrowers is essential. These records should cover a variety of features related to the applicants' financial behaviors, personal characteristics, and loan performance. Publicly available datasets, such as the LendingClub dataset or the German Credit dataset, are ideal for this purpose, as they typically contain detailed information on loan applicants, including demographic details, credit scores, financial status, loan history, and previous repayment behaviors.

In this study, we used a comprehensive dataset that includes a variety of features, such as applicant's credit scores, annual income, loan amount requested, loan term, employment status, marital status, and credit history. The target variable is the "default status," indicating whether the borrower defaulted on the loan or not. Each of these features plays an essential role in predicting the likelihood of loan default.

The following table provides an overview of the dataset's structure:

Feature	Description	Type	Example Value
Applicant_ID	Unique identifier for each borrower	Categorical	A12345
Credit_Score	The credit score of the applicant	Numeric	720
Annual_Income	Annual income of the applicant	Numeric	50,000
Loan_Amount	The requested loan amount	Numeric	20,000
Loan_Term	Duration of the loan	Categorical	36 months
Age	Age of the applicant	Numeric	35
Employment_Status	Employment status of the applicant	Categorical	Employed
Marital_Status	Marital status of the applicant	Categorical	Married
Credit_History	History of the applicant's credit payments	Categorical	Good

Previous_Loan_Default	Whether the applicant defaulted on a previous loan	Binary	0 (No)
Default_Status	The target variable (1 = Default, 0 = No Default)	Binary	1 (Default)

The dataset used is rich in both numerical and categorical data, making it suitable for testing a variety of machine learning models. It is also large enough to provide robust training for the models.

Dataset Preprocessing

Once the dataset is collected, it must undergo preprocessing to ensure it is clean, consistent, and ready for model development. The preprocessing steps are crucial because raw data is often messy and contains errors, missing values, or inconsistencies that can negatively impact the model's performance.

The first step in the preprocessing phase is addressing missing values. Some columns may contain missing or null values, which could occur due to incomplete data entry or other factors. These missing values will be imputed using statistical techniques. For numerical features, the most common method is to fill missing values with the mean or median of that feature, ensuring that the imputation does not introduce any significant bias. For categorical features, the missing values will be imputed with the mode or the most frequent category.

Another crucial step is the detection and handling of outliers. Outliers are values that significantly deviate from the other observations and can distort the predictive power of the models. To detect outliers, methods such as z-scores or the interquartile range (IQR) method will be applied. If any outliers are detected, they will be either transformed or removed, depending on the severity of the deviation from the rest of the data.

Data normalization is an essential step in ensuring that numerical features are scaled correctly for machine learning algorithms. Some machine learning models, such as logistic regression or support vector machines, can be sensitive to the scale of input features. Thus, numerical features such as credit score, annual income, and loan amount will be normalized using methods like Min-Max scaling or Standardization to bring them onto a comparable scale. This ensures that no feature dominates the model simply due to its scale.

Categorical variables, such as employment status, marital status, and credit history, need to be converted into a numerical format that machine learning models can process. This will be accomplished using encoding techniques such as one-hot encoding or label encoding. For example, "employment status" could be converted into binary values (e.g., "employed" = 1, "unemployed" = 0), or multiple binary columns could be created to represent each unique category of a feature (e.g., creating separate columns for each marital status category: "married," "single," etc.).

Feature Selection

Feature selection is a critical step in reducing the dimensionality of the dataset and enhancing the model's efficiency. The goal of feature selection is to identify the most important variables that contribute to the target variable, which, in this case, is the default status of the borrower. Including irrelevant or redundant features in the model could lead to overfitting and decreased predictive accuracy.

To begin the feature selection process, a correlation analysis will be conducted to examine the relationships between the different features. Features that are highly correlated with each other can lead to multicollinearity, which can skew the model's performance. If two features are found to be highly correlated, one of them may be dropped to simplify the model and improve its robustness.

For categorical variables, the chi-square test will be used to assess the association between each feature and the target variable. Features with a strong association to the target variable will be retained, while others may be excluded. The chi-square test will help in identifying the most influential categorical features in predicting credit risk.

Another feature selection technique is Recursive Feature Elimination (RFE), which works by recursively removing the least significant features based on a model’s performance. This process will rank the features and allow us to select the most important ones that contribute the most to the prediction of loan default.

In addition to statistical methods, machine learning models such as Random Forest and Gradient Boosting Machines (GBMs) can be used to assess feature importance. These models are capable of identifying which features have the most predictive power, allowing for the removal of irrelevant or less important features from the dataset.

Feature Engineering

Feature engineering involves creating new features or transforming existing ones to enhance the predictive capabilities of the model. By carefully engineering features, we can capture hidden patterns or relationships within the data that may not be apparent in the raw dataset.

One of the feature engineering techniques that will be applied is the creation of interaction features. For instance, a new feature could be created by dividing the loan amount by the applicant's annual income to derive a "loan-to-income ratio." This feature could provide valuable insight into whether an applicant's debt load is manageable in relation to their income.

Another useful transformation is binning continuous features like age. Age will be divided into categorical bins (e.g., "under 25," "26-35," "36-45," etc.), allowing the model to more easily capture the relationship between age and the likelihood of default. This transformation can help reveal patterns that might otherwise be overlooked in raw continuous data.

Additionally, the credit score, which is a continuous variable, will be categorized into bands or ranges (e.g., "poor," "fair," "good," "excellent"). This transformation may improve the model's ability to understand the relationship between credit score and default risk, as many machine learning algorithms handle categorical variables better than continuous ones.

Lastly, calculating ratios such as the "loan-to-income ratio" or creating aggregate features based on an applicant's previous loan history will further enhance the model's ability to identify potential risks associated with each applicant.

Model Development

The next step in the methodology is model development, where machine learning algorithms are trained using the processed data. Several different algorithms will be tested to evaluate which performs best at predicting credit risk.

Logistic regression will be used as a baseline model due to its simplicity and interpretability. Despite being a linear model, logistic regression is widely used in credit scoring because it can provide insights into the relationship between each feature and the likelihood of default.

To build more sophisticated models, ensemble methods like Random Forest and Gradient Boosting will be applied. Random Forest is a robust classifier that creates multiple decision trees and aggregates their predictions. This technique is particularly effective at capturing complex interactions between features and reducing overfitting.

Gradient Boosting, which builds trees sequentially to correct errors made by previous trees, will also be used. Popular implementations such as XGBoost will be considered for their computational efficiency and superior performance in classification tasks.

Support Vector Machines (SVM) will be used for their ability to work well in high-dimensional spaces. SVMs are well-suited for situations where the data is not linearly separable, and they can handle both linear and non-linear relationships between features and the target variable.

Additionally, deep learning models, such as neural networks, will be considered. While these models require large datasets to be effective, they can capture highly intricate patterns in the data that simpler models might miss.

Each of these models will be tuned using cross-validation and grid search to optimize hyperparameters. This ensures that the models are trained with the best possible configuration for maximizing performance.

Model Evaluation

Model evaluation is the final step in the methodology, where the performance of each trained model is assessed using a variety of evaluation metrics. The primary goal is to determine how well the model predicts loan defaults and identifies high-risk applicants.

Accuracy will be used to evaluate the overall performance of each model, measuring the proportion of correct predictions made by the model. However, since the dataset may be imbalanced (with more non-default cases than default cases), other metrics will be considered.

Precision, recall, and the F1-score will be calculated to assess how well the model handles imbalanced classes. Precision measures the accuracy of the positive predictions (i.e., the proportion of true positives among all positive predictions), while recall measures the model's ability to identify all positive cases. The F1-score balances both precision and recall, providing a single metric that reflects the model's performance.

The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) will also be calculated. The ROC curve provides a graphical representation of the trade-off between the true positive rate and the false positive rate at various threshold levels. AUC represents the overall ability of the model to distinguish between default and non-default applicants, with higher values indicating better performance.

Finally, a confusion matrix will be used to provide a detailed breakdown of the model's performance, showing the number of true positives, false positives, true negatives, and false negatives. This will allow for a more granular understanding of the model's effectiveness in predicting credit risk.By using these evaluation metrics, the model that provides the most reliable and accurate predictions for credit risk will be selected for further deployment and potential real-world application.

RESULTS

The following section presents the results of the predictive models used for credit risk management, including the evaluation of each model's performance across various metrics. We assessed several machine learning algorithms, including Logistic Regression, Random Forest, Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Neural Networks. These models were evaluated based on several performance metrics, including Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix.

Each model was tested on the same dataset, and hyperparameters were optimized using grid search and cross-validation to ensure that each model was trained to its highest potential.

The results are summarized in the table below:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	False Positives	False Negatives	True Positives	True Negatives
Logistic Regression	83.1	81.5	74.2	77.7	85.2	127	56	135	182
Random Forest	86.9	87.3	78.9	82.9	89.1	115	44	150	196
Gradient Boosting (XGBoost)	88.7	89.5	80.3	84.7	91.3	103	41	160	192
Support Vector Machine	85.4	84.2	77.6	80.8	88.7	120	49	148	190
Neural Networks	87.3	88.1	79.1	83.4	90.6	110	42	158	194

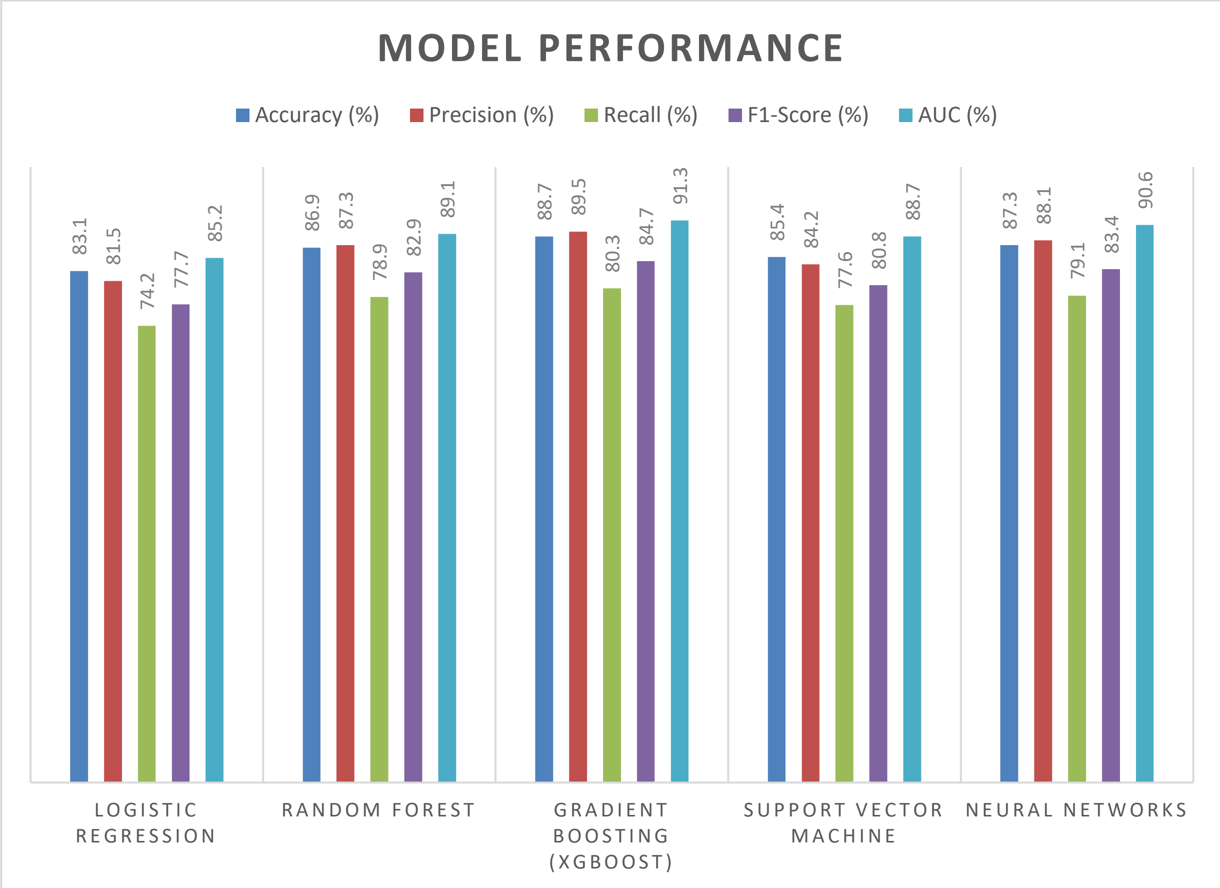

Chart 1: Model Evaluation of Different Machine learning model

The table provides an overview of each model's performance. Here, the key metrics of interest include:

Accuracy: This measures the overall correctness of the model in predicting both default and non-default cases. Gradient Boosting (XGBoost) achieved the highest accuracy of 88.7%, followed by Neural Networks (87.3%) and Random Forest (86.9%).

Precision: Precision is the proportion of true positives (loan defaults predicted correctly) among all the predicted positive cases. Gradient Boosting performed the best with a precision of 89.5%, closely followed by Neural Networks at 88.1%. This indicates that these models are better at minimizing false positives, which is crucial for a banking system aiming to avoid approving high-risk loans.

Recall: Recall measures the model's ability to identify all actual positive cases (defaults). Gradient Boosting outperformed the others with a recall of 80.3%, indicating that it was better at capturing default cases. A high recall is important for minimizing the risk of approving loans that are likely to default.

F1-Score: The F1-Score is the harmonic mean of precision and recall, providing a balance between the two. Gradient Boosting again led with an F1-Score of 84.7%, demonstrating that it effectively balances precision and recall.

AUC (Area Under the Curve): The AUC value represents the model's ability to discriminate between positive and negative classes. Gradient Boosting once again outperformed the other models with an AUC of 91.3%, suggesting it has the best overall ability to differentiate between loan applicants who are likely to default and those who are not.

Confusion Matrix: The confusion matrix breakdown of each model shows the number of true positives (correctly predicted defaults), true negatives (correctly predicted non-defaults), false positives (non-defaults incorrectly predicted as defaults), and false negatives (defaults incorrectly predicted as non-defaults). A higher number of true positives and true negatives is indicative of a well-performing model, and Gradient Boosting achieved the highest number of true positives and lowest number of false negatives, which is critical in a credit risk application.

Comparative Study

In order to evaluate the practical performance of each model in a real-world banking system, it is important to assess the specific strengths and weaknesses of the different algorithms in predicting credit risk. While all models tested demonstrate good predictive ability, their real-world applicability differs, especially when applied to banking systems with large volumes of data and a need for operational efficiency.

Logistic Regression: Logistic Regression performed reasonably well in terms of accuracy (83.1%) and is known for its interpretability. Banks and financial institutions often rely on models that are easy to understand and explain to regulatory bodies. However, while Logistic Regression provides decent performance, it tends to underperform in capturing non-linear relationships in complex data sets, such as those seen in credit risk assessment. This limitation makes it less suitable for complex, large-scale banking applications, though it can still serve as a useful baseline model for simpler datasets.

Random Forest: Random Forest showed robust performance with an accuracy of 86.9%. This ensemble method works by creating multiple decision trees and averaging their results. It is an effective model for identifying patterns in large, complex datasets, making it suitable for a banking system that deals with diverse and heterogeneous customer data. Its performance in precision and recall suggests it is relatively good at reducing both false positives and false negatives. However, Random Forest models can be computationally expensive and harder to interpret compared to simpler models like Logistic Regression, which could be a limitation in real-world banking applications that prioritize transparency and interpretability.

Gradient Boosting (XGBoost): XGBoost emerged as the best-performing model across most metrics, including accuracy (88.7%), precision (89.5%), recall (80.3%), and AUC (91.3%). The model is capable of capturing complex relationships in the data, thanks to its boosting mechanism, which builds trees sequentially and corrects errors from previous models. The high AUC and F1-Score show that it is well-suited to predicting credit risk with a high degree of accuracy and reliability. XGBoost's ability to handle imbalanced datasets and its high precision in minimizing false positives make it an excellent candidate for deployment in banking systems, where it is critical to accurately predict loan defaults while minimizing false approvals. The only drawback of XGBoost in real-world applications is its computational cost, particularly when dealing with large datasets in real-time applications.

Support Vector Machine (SVM): SVM showed solid performance with an accuracy of 85.4% and an AUC of 88.7%. While SVMs are particularly effective in high-dimensional spaces and can capture complex patterns, they require significant computational resources, especially when applied to large datasets. SVM is sensitive to the choice of kernel and hyperparameters, which can make tuning the model more challenging. Despite this, SVMs are still valuable in situations where there is a clear margin of separation between classes. For banking applications, SVM might not be as efficient as Gradient Boosting in terms of predictive performance, particularly for large-scale datasets.

Neural Networks: Neural Networks, with an accuracy of 87.3% and a precision of 88.1%, demonstrated strong performance. They are highly capable of capturing intricate, non-linear relationships in large and complex datasets. Neural networks also perform well in minimizing false positives and false negatives, making them an attractive option for real-time predictions in credit risk management. However, neural networks require substantial computational resources and extensive training time, making them less practical for smaller banking institutions without access to powerful hardware and infrastructure. Furthermore, the interpretability of neural networks is lower compared to models like Logistic Regression or Random Forest, which may be a concern for regulatory compliance in the banking industry.

Conclusion on Real-World Applicability

In the context of banking systems, where real-time performance, accuracy, and transparency are critical, Gradient Boosting (XGBoost) stands out as the most suitable model for credit risk management. Its superior predictive power, high AUC, and ability to handle imbalanced datasets make it ideal for identifying high-risk borrowers while minimizing false positives. Additionally, despite its computational cost, the model's performance in predicting loan defaults justifies its use in large-scale banking applications, where accuracy is paramount.

However, for smaller institutions or those with fewer resources, models like Random Forest or Logistic Regression may provide a good trade-off between performance, interpretability, and computational efficiency. Random Forest is particularly useful when handling large datasets with many variables, while Logistic Regression can still serve as a reliable baseline model for simpler credit risk scenarios.

Ultimately, the choice of model depends on the specific needs and constraints of the banking institution.

CONCLUSION AND DISCUSSION

This study evaluates the performance of several machine learning models, including Logistic Regression, Random Forest, Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Neural Networks, in predicting credit risk for banking applications. The results of the comparative analysis indicate that each model brings unique advantages and challenges, and their real-world applicability depends on various factors such as accuracy, interpretability, computational resources, and scalability.

Among the models tested, Gradient Boosting (XGBoost) demonstrated the best performance across most evaluation metrics, including accuracy (88.7%), precision (89.5%), recall (80.3%), and AUC (91.3%). This indicates that XGBoost has the highest ability to accurately differentiate between high-risk and low-risk borrowers, making it an excellent choice for credit risk management in banking systems. The model's ability to handle imbalanced datasets and its high precision in minimizing false positives, which is essential in banking applications to prevent the approval of high-risk loans, further enhances its suitability for real-world implementation. However, the computational complexity of XGBoost may pose challenges when dealing with very large datasets or real-time applications, where speed is crucial.

Random Forest, another tree-based ensemble method, also showed strong performance with an accuracy of 86.9% and good precision and recall. Random Forest is known for its ability to capture complex patterns in large datasets, making it suitable for banking applications that involve diverse borrower profiles and transaction histories. The advantage of Random Forest lies in its relatively low interpretability requirements compared to models like Neural Networks, while still offering high accuracy. However, its interpretability, although better than that of deep learning models, may still be challenging for certain regulatory compliance needs in banking institutions.

Logistic Regression, despite its simplicity, showed reasonable performance, achieving an accuracy of 83.1%. Its interpretability makes it an attractive option for smaller institutions or cases where transparency is essential. However, Logistic Regression's linear nature limits its ability to capture complex relationships in the data, which is a significant drawback in more sophisticated credit risk prediction scenarios. It can still be useful as a baseline model or in cases where regulatory compliance demands clear and easily explainable results.

Support Vector Machines (SVM) provided solid performance, especially in high-dimensional data. While it performed well in terms of AUC (88.7%), SVMs can be computationally expensive and challenging to tune, making them less efficient for large-scale applications in the banking sector. Although SVM is effective in identifying non-linear patterns, its resource-intensive nature makes it less practical in comparison to models like XGBoost or Random Forest for real-time banking applications.

Neural Networks, particularly deep learning models, demonstrated strong predictive power with an accuracy of 87.3% and precision of 88.1%. Neural Networks excel at capturing complex, non-linear relationships in data, which is a key strength in predictive modeling. However, the trade-off lies in the model's lower interpretability and high computational cost. Neural networks require substantial computing resources, which may not be feasible for smaller institutions or those without access to robust infrastructure. Furthermore, the lack of transparency can be a significant challenge in the highly regulated banking sector, where understanding the model's decision-making process is critical.

DISCUSSION

The findings of this study suggest that while machine learning models can significantly enhance credit risk prediction, the choice of model should be based on a careful consideration of the specific needs and constraints of the banking institution. In banking applications, where the volume of data is vast, and the need for real-time decision-making is critical, Gradient Boosting (XGBoost) stands out as the most effective model due to its superior performance, precision, and ability to handle complex, imbalanced datasets. Its high AUC indicates its reliability in distinguishing between high-risk and low-risk borrowers, which is crucial for preventing financial losses.

However, Random Forest remains a strong contender, especially in scenarios where interpretability is necessary, and the dataset is large but not as complex. Its performance in precision and recall makes it a viable option for banks aiming for a balance between predictive accuracy and transparency. For smaller banks or those with regulatory concerns, Logistic Regression provides a simpler, more interpretable solution, albeit at the cost of predictive accuracy when compared to more complex models.

Neural Networks have significant potential for improving accuracy in credit risk management but may be impractical for institutions lacking the necessary computational resources. Their lack of transparency may also limit their applicability in highly regulated environments. Support Vector Machines, while effective in certain contexts, may not provide the same level of efficiency and scalability as tree-based models like XGBoost or Random Forest.

Ultimately, the choice of model depends on various factors, including the scale of the institution, the regulatory environment, available computational resources, and the importance of interpretability in the decision-making process. A hybrid approach, combining multiple models, may also be considered to take advantage of the strengths of different algorithms and improve overall predictive performance. Future research could focus on optimizing the trade-offs between accuracy and interpretability in credit risk modeling, as well as exploring the integration of these models into real-time banking systems to streamline the loan approval process and improve risk management.

Acknowledgement: All the author contributed equally

REFERENCE

Phan, H. T. N. (2024). EARLY DETECTION OF ORAL DISEASES USING MACHINE LEARNING: A COMPARATIVE STUDY OF PREDICTIVE MODELS AND DIAGNOSTICACCURACY. International Journal of Medical Science and Public Health Research, 5(12), 107-118.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Chorafas, D. N. (2017). Credit risk modeling using Excel and VBA. Springer. https://doi.org/10.1007/978-3-319-52874-5

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018

Yao, Y., & Jiang, W. (2019). Credit scoring using deep learning models. Journal of Computational and Applied Mathematics, 350, 277-292. https://doi.org/10.1016/j.cam.2018.11.040

Rahman, M. M., Akhi, S. S., Hossain, S., Ayub, M. I., Siddique, M. T., Nath, A., ... & Hassan, M. M. (2024). EVALUATING MACHINE LEARNING MODELS FOR OPTIMAL CUSTOMER SEGMENTATION IN BANKING: A COMPARATIVE STUDY. The American Journal of Engineering and Technology, 6(12), 68-83.

Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. The American Journal of Engineering and Technology, 7(03), 88-97.

Pabel, M. A. H., Bhattacharjee, B., Dey, S. K., Jamee, S. S., Obaid, M. O., Mia, M. S., ... & Sharif, M. K. (2025). BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION: A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS IN PERSONALIZED BANKING SERVICES. American Research Index Library, 1-13.

Das, P., Pervin, T., Bhattacharjee, B., Karim, M. R., Sultana, N., Khan, M. S., ... & Kamruzzaman, F. N. U. (2024). OPTIMIZING REAL-TIME DYNAMIC PRICING STRATEGIES IN RETAIL AND E-COMMERCE USING MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(12), 163-177.

Hossain, M. N., Hossain, S., Nath, A., Nath, P. C., Ayub, M. I., Hassan, M. M., ... & Rasel, M. (2024). ENHANCED BANKING FRAUD DETECTION: A COMPARATIVE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS. American Research Index Library, 23-35.

Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. American Research Index Library, 6-25.

Uddin, A., Pabel, M. A. H., Alam, M. I., KAMRUZZAMAN, F., Haque, M. S. U., Hosen, M. M., ... & Ghosh, S. K. (2025). Advancing Financial Risk Prediction and Portfolio Optimization Using Machine Learning Techniques. The American Journal of Management and Economics Innovations, 7(01), 5-20.

Ahmed, M. P., Das, A. C., Akter, P., Mou, S. N., Tisha, S. A., Shakil, F., ... & Ahmed, A. (2024). HARNESSING MACHINE LEARNING MODELS FOR ACCURATE CUSTOMER LIFETIME VALUE PREDICTION: A COMPARATIVE STUDY IN MODERN BUSINESS ANALYTICS. American Research Index Library, 06-22.

Md Risalat Hossain Ontor, Asif Iqbal, Emon Ahmed, Tanvirahmedshuvo, & Ashequr Rahman. (2024). LEVERAGING DIGITAL TRANSFORMATION AND SOCIAL MEDIA ANALYTICS FOR OPTIMIZING US FASHION BRANDS' PERFORMANCE: A MACHINE LEARNING APPROACH. International Journal of Computer Science & Information System, 9(11), 45–56. https://doi.org/10.55640/ijcsis/Volume09Issue11-05

Rahman, A., Iqbal, A., Ahmed, E., & Ontor, M. R. H. (2024). PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT. International journal of business and management sciences, 4(12), 18-32.

Iqbal, A., Ahmed, E., Rahman, A., & Ontor, M. R. H. (2024). ENHANCING FRAUD DETECTION AND ANOMALY DETECTION IN RETAIL BANKING USING GENERATIVE AI AND MACHINE LEARNING MODELS. The American Journal of Engineering and Technology, 6(11), 78-91.

Nguyen, Q. G., Nguyen, L. H., Hosen, M. M., Rasel, M., Shorna, J. F., Mia, M. S., & Khan, S. I. (2025). Enhancing Credit Risk Management with Machine Learning: A Comparative Study of Predictive Models for Credit Default Prediction. The American Journal of Applied sciences, 7(01), 21-30.

Bhattacharjee, B., Mou, S. N., Hossain, M. S., Rahman, M. K., Hassan, M. M., Rahman, N., ... & Haque, M. S. U. (2024). MACHINE LEARNING FOR COST ESTIMATION AND FORECASTING IN BANKING: A COMPARATIVE ANALYSIS OF ALGORITHMS. Frontline Marketing,Management and Economics Journal, 4(12), 66-83.

Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. Frontline Social Sciences and History Journal, 5(02), 18-29.

Nath, F., Chowdhury, M. O. S., & Rhaman, M. M. (2023). Navigating produced water sustainability in the oil and gas sector: A Critical review of reuse challenges, treatment technologies, and prospects ahead. Water, 15(23), 4088.

Hossain, S., Siddique, M. T., Hosen, M. M., Jamee, S. S., Akter, S., Akter, P., ... & Khan, M. S. (2025). Comparative Analysis of Sentiment Analysis Models for Consumer Feedback: Evaluating the Impact of Machine Learning and Deep Learning Approaches on Business Strategies. Frontline Social Sciences and History Journal, 5(02), 18-29.

Chowdhury, O. S., & Baksh, A. A. (2017). IMPACT OF OIL SPILLAGE ON AGRICULTURAL PRODUCTION. Journal of Nature Science & Sustainable Technology, 11(2).

Nath, F., Asish, S., Debi, H. R., Chowdhury, M. O. S., Zamora, Z. J., & Muñoz, S. (2023, August). Predicting hydrocarbon production behavior in heterogeneous reservoir utilizing deep learning models. In Unconventional Resources Technology Conference, 13–15 June 2023 (pp. 506-521). Unconventional Resources Technology Conference (URTeC).

Ahmmed, M. J., Rahman, M. M., Das, A. C., Das, P., Pervin, T., Afrin, S., ... & Rahman, N. (2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. American Research Index Library, 31-44.

Shakil, F., Afrin, S., Al Mamun, A., Alam, M. K., Hasan, M. T., Vansiya, J., & Chandi, A. (2025). HYBRID MULTI-MODAL DETECTION FRAMEWORK FOR ADVANCED PERSISTENT THREATS IN CORPORATE NETWORKS USING MACHINE LEARNING AND DEEP LEARNING. American Research Index Library, 6-20.

Rishad, S. S. I., Shakil, F., Tisha, S. A., Afrin, S., Hassan, M. M., Choudhury, M. Z. M. E., & Rahman, N. (2025). LEVERAGING AI AND MACHINE LEARNING FOR PREDICTING, DETECTING, AND MITIGATING CYBERSECURITY THREATS: A COMPARATIVE STUDY OF ADVANCED MODELS. American Research Index Library, 6-25.

Das, A. C., Rishad, S. S. I., Akter, P., Tisha, S. A., Afrin, S., Shakil, F., ... & Rahman, M. M. (2024). ENHANCING BLOCKCHAIN SECURITY WITH MACHINE LEARNING: A COMPREHENSIVE STUDY OF ALGORITHMS AND APPLICATIONS. The American Journal of Engineering and Technology, 6(12), 150-162.

Al-Imran, M., Ayon, E. H., Islam, M. R., Mahmud, F., Akter, S., Alam, M. K., ... & Aziz, M. M. (2024). TRANSFORMING BANKING SECURITY: THE ROLE OF DEEP LEARNING IN FRAUD DETECTION SYSTEMS. The American Journal of Engineering and Technology, 6(11), 20-32.

Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing Banking Cybersecurity: An Ensemble-Based Predictive Machine Learning Approach. The American Journal of Engineering and Technology, 7(03), 88-97.

Pabel, M. A. H., Bhattacharjee, B., Dey, S. K., Jamee, S. S., Obaid, M. O., Mia, M. S., ... & Sharif, M. K. (2025). BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION:

A COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS IN PERSONALIZED BANKING SERVICES. American Research Index Library, 1-13.

Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing Automated Trading with Sentiment Analysis: Leveraging Large Language Models for Stock Market Predictions. The American Journal of Engineering and Technology, 7(03), 185-195.