International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-14 Issue-10, September 2025

Identifying Optimal Innovative Machine Learning Models for Predicting U.S. Mortgage Loan Defaults: A Comparative Analysis

Chi Ton Cong Nguyen, Victoria N. Dean

Abstract*: Mortgage loan defaults pose substantial risks to the financial industry, and accurately predicting these defaults remains a challenge. Traditional credit scoring models often lack accuracy and computational efficiency because they are not data-driven and fail to capture special data patterns that drive borrower default behavior. This limitation becomes especially significant in volatile housing markets, where early detection of defaults can substantially reduce credit losses compared to identifying them at a later stage. Timely and accurate prediction of mortgage loan defaults plays a vital role in formulating effective credit risk management strategies. This research implemented a framework to address this challenge by leveraging cutting-edge Artificial Intelligence / Machine Learning (AI/ML) algorithms and conducting a comparative analysis of model performance to identify the optimal model for predicting mortgage loan defaults. The proposed framework systematically trained and evaluated each model on an extensive dataset comprising over 100,000 loans, featuring a rich set of loan, borrower, and property characteristics from Freddie Mac, as well as macroeconomic factors. Model performance was evaluated using key metrics including accuracy, AUC, F1 scores, and ROC curves. The paper discovered that Extreme Gradient Boosting (XGBoost) was the top performer, offering superior performance and robustness to overfitting, compared to other ML models, including Logistic Regression, Neural Network, Decision Tree, Gradient Boosting, and Random Forest. The results demonstrated that XGBoost achieved the best performance across all evaluation metrics, with 99% accuracy on the training data, 98% on the testing data, and more than 90% for all other metrics. The robust predictive power of XGBoost is mainly due to its ensemble and regularisation techniques, which minimise errors and the overfitting problem simultaneously. These findings contribute a crucial benchmark for mortgage default modeling practice and develop an innovative financial ML technique, XGBoost, for predicting "good" and "bad" loans accurately. Given XGBoost's exceptional performance in predicting mortgage loan defaults, the study offers an innovative AI solution for credit risk assessment and smart lending decisions for banks, mortgage lenders, and financial institutions in the FinTech industry. The research also highlights a superior AI/ML classification algorithm for any field.

***Keywords*: Artificial Intelligence/Machine Learning, Credit Scoring Models, FinTech Industry, Financial Innovation, Mortgage Loan Defaults.**

Manuscript received on 30 July 2025 | First Revised Manuscript received on 06 August 2025 | Second Revised Manuscript received on 18 August 2025 | Manuscript Accepted on 15 September 2025 | Manuscript published on 30 September 2025.

*Correspondence Author(s)

Chi Ton Cong Nguyen*, James Madison High School, Vienna, VA, USA. Email ID: chiton0408@gmail.com, ORCID ID: 0009-0002-8065-7215

Dr. Victoria N. Dean, AI/ML Principal Researcher, New Jersey, USA. Email ID: victoria.usacademy@gmail.com, ORCID ID: 0009-0006-0629-3416

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Abbreviations:

ACC: Accuracy
AI: Artificial Intelligence
AUC: Area Under the Curve
CLTV: Current Loan to Value
CUPB: Current Unpaid Balance
DT: Decision Tree
FICO: Credit Score
FN: False Negative
FP: False Positive
GB: Gradient Boosting
LR: Logistic Regression
ML: Machine Learning
MLP: Multilayer Perceptrons
NN: Neural Network
RF: Random Forest
ROC: Receiver Operating Characteristic
TN: True Negative
TP: True Positive
XGBoost: Extreme Gradient Boosting

I. INTRODUCTION

Mortgage lending is one of the most significant segments in the FinTech industry as it represents a high-value and long-term financial product in the consumer finance sector. Mortgage lending presents significant opportunities for investment but also carries substantial risks. To manage credit risks effectively, credit scoring models were widely used to determine borrowers' creditworthiness and make informed lending decisions [1]. Credit scoring models used to detect mortgage loan defaults at an early stage and mitigate the loss in the long run are crucial because widespread mortgage defaults, especially subprime mortgages, were major contributors to the 2008 financial crisis, economic disruption, and global turmoil [2]. As a result, developing accurate predictive models of mortgage loan defaults is necessary for effective risk management in the financial industry and for sustaining macroeconomic and financial stability at both national and international levels.

Due to the increasing availability of consumer finance data and the outstanding achievements of state-of-the-art artificial intelligence (AI), machine learning (ML) classification algorithms have been widely used in the financial industry [3]. AI/ML models are innovative techniques that offer advanced capabilities for analyzing complex and large datasets involving borrower behavior, loan features, and risk profiles. The ability to identify special data patterns enables AI/ML models to improve mortgage loan default predictions significantly.

Given the rapid digitalisation, the increasing availability of financial data,

© Copyright: All rights reserved.

and the growing need for improved risk management of mortgage loans in the finance sector, this study experimented with leveraging cutting-edge AI/ML models and identifying the optimal ML model for predicting mortgage loan defaults. The best ML model plays a vital role in detecting loan defaults accurately, providing early warning indicators for financial risk intervention, reducing credit risks, and mitigating financial losses for practical use in the finance industry.

The paper is structured in the following sections. Section 2 describes the data of mortgage loans and discusses the methodology of various ML algorithms used in this study. Section 3 illustrates the model development process, including data preprocessing, sampling, feature selection, and model evaluation. Section 4 provides a summary and discussion of analytical results across ML models. The last section concludes the main finding of the comparative analysis.

II. DATA AND METHODS

A. Data

The data for this research contains 103,599 mortgage loans from Freddie Mac data in 2020 [4]. Each observation corresponds to a unique loan with one specific loan status and twenty features. The outcome variable is the mortgage loan status, indicating whether the loan was defaulted (i.e., missed payments for at least six months). The data consists of 10,867 instances of defaulted mortgage loans and 92,732 cases of non-defaulted loans. The target variable is binary. Each loan includes a rich set of loan, borrower, and property characteristics, such as credit score, current loan-to-value ratio, origination interest rate, loan age, current unpaid balance, type of property (single-family home, condo), and whether the borrower is a first-time home buyer (see Table 1 for summary statistics). Additionally, the data also covers macroeconomic features (i.e., unemployment rates and inflation rates) based on borrowers' geographical locations.

Table-I: Summary Statistics (N =103,599)

Features	Feature Description	Mean
FICO	Credit Score	742.27
CLTV	Current Loan to Value Ratio	0.83
CUPB	Current Unpaid Balance	267,319
ORG_RATE	Origination Interest Rate	3.39
LOAN_AGE	Loan Age	14.77
DRATE	Change in Interest Rate	-0.23
Purchase	Purchase	0.39
NonCORefi	No Cash-out Refinance	0.4
CORefi	Cash-out Refinance	0.21
SF	Single-family home	0.72
PUD	Planned Unit Developments	0.2
Otherprop	Other property types	0.08
Investment	Investment	0.05
Owner_Occupied	Owner occupied	0.92
Second_home	Second home	0.03
FTHB	First-time home buyer	0.18
Outcome		
DEFAULT	Missing payments for 6+ months	0.105

B. Machine Learning Models

In this study, a performance comparison is conducted across several AI/ML models to identify which ML model has the best accuracy. The models examined in this work include Logistic Regression (also known as the logit model), Decision Trees, Neural Networks, Gradient Boosting, Random Forests, and XGBoost.

i. Logistic Regression (LR)

LR or Logit Model has been considered a classic statistical model in credit risk assessment [3]. LR is widely used for binary classification problems due to its simple parametric structure (see equation (1)), ease of interpretation, and understanding how various borrower, loan, and property characteristics influence the probability of loan defaults [5]. The LR can be expressed as a linear regression, where the dependent variable is the log-odds of loan defaults, and the independent variables are borrower, loan, and property characteristics (X) (see equation (2)).

$$P(Y_i = 1) = \frac{exp\ (X\beta)}{1+exp\ (X\beta)} \quad \ldots \quad (1)$$

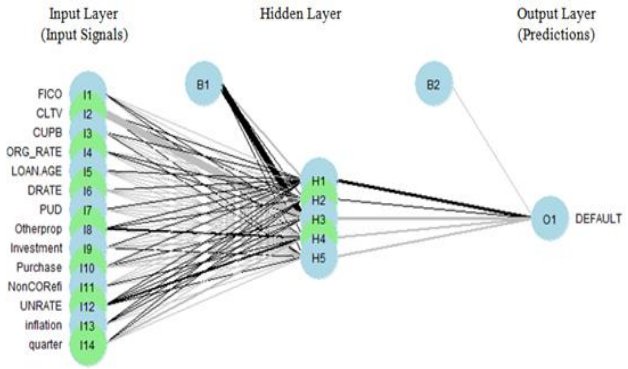
$$Log\left(\frac{P(Y_i=1)}{1-P(Y_i=1)}\right) = X\beta \quad \ldots \quad (2)$$

where X is a matrix containing borrower, loan, and property characteristics such as FICO, CLTV, first-time home buyer, condo property, and βs represent the coefficients of those predictors.

Despite the straightforward interpretation of the LR, some drawbacks of the linearity and independence assumptions still need to be addressed. Other cutting-edge ML algorithms can offer higher predictive power than LR because they provide a more flexible and robust framework for modelling nonlinear relationships in data with complex behaviour of various borrower, loan, and property features. This makes ML models innovative tools in credit decision-making, providing enhanced predictive accuracy, deeper insights, and greater practical use.

ii. Neural Network (NN) Algorithm

NN resembles a biological structure with a network of neurons similar to the human brain, with multilayer perceptrons (MLPs) and a learning network to process data, recognize patterns, make decisions, and transform inputs to outputs [6]. MLP presents a network of interconnected neurons, comprising multiple layers: an input layer, one or more hidden layers, and an output layer (see Figure 1). Since the network uses nonlinear activation functions, it helps the NN algorithm learn more complex data patterns and transform the input signals to final outputs with predictions in a nonlinear manner [6]. NN has been demonstrated as a stronger candidate for credit scoring models than traditional statistical methods, including logistic regressions [7].

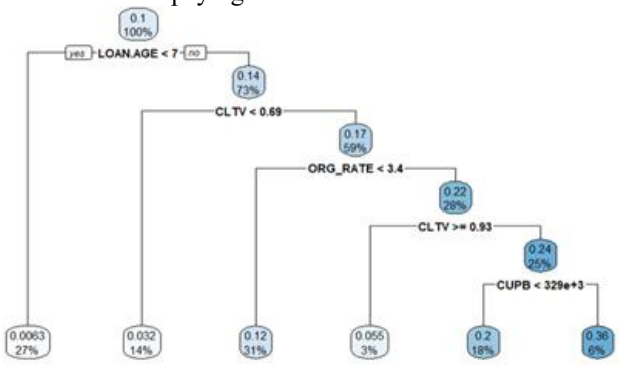


[Fig.1: Diagram of Neural Network Algorithm with Multiple Layers]

© Copyright: All rights reserved.

iii. Decision Tree (DT) Algorithm

DT was designed as a tree structure that answers a series of "Yes/No" questions, mimicking a human decision-making process [8]. It functions like a hierarchical flowchart, making it easy to interpret and visualise how each decision was made based on the cutoff of features in the data (see Figure 2). Based on the importance of the decision and its associated feature, the data was recursively split into two parts, from the root to each branch and each leaf node [8]. For example, the tree or the branch was divided into two parts based on Loan age, CLTV, origination interest rates, and CUPB (see Figure 2). The split was conducted by minimizing Gini impurity or entropy in the remaining data after the split [8]. This splitting process was implemented until the decision tree was fully grown. From a credit risk standpoint, the DT model identifies the most effective feature to distinguish between defaults and non-defaults, and repeats this division process until all essential features are utilised to detect and classify default behaviour from repaying behaviour.



[Fig.2: Diagram of Decision Tree Model with Feature Splits]

iv. Random Forest (RF) Algorithm

RF is a robust supervised ML algorithm that collects multiple DTs to enhance accuracy and reliability for classification purposes by averaging the predictions from its inclusive decision trees [9]. By building various trees through multiple sampling, variance can be reduced and prediction can be improved, given the rising cost of training time. Similar to the DT algorithm, key variables that play an effective role in classifications were used more frequently in RF.

RF is an ensemble method that has a better performance than a single decision tree [3]. This method generally demonstrates high predictive power, a relatively simple implementation, and a strong capacity to identify and model non-linear, special, and high-dimensional data patterns, as well as feature importance. Due to these advantages, RF is an effective ML method that is widely used in finance [10] and housing market [11].

v. Gradient Boosting (GB) Algorithm

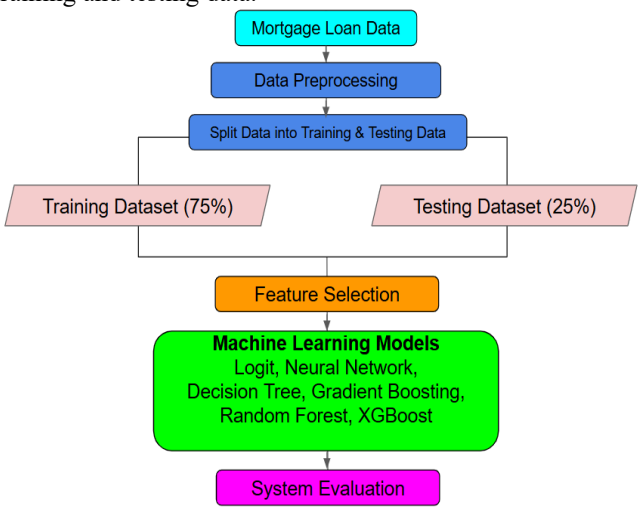
GB is an ensemble technique that combines "weak learners" to fix their mistakes and to build a strong model [12]. GB is a tree-based technique; however, it learns from the misclassification of the previous tree and trains a new model to correct that error by minimising the loss function with an adaptive learning rate. This process is repeated over multiple iterations to reduce bias, with a newly trained model added at each stage, until the final model achieves the highest accuracy. GB is a robust algorithm for classification tasks; thus, it is a potential candidate for classifying mortgage loan defaults.

vi. Extreme Gradient Boosting (XGBoost) Algorithm

Extreme Gradient Boosting (XGBoost) is an ML algorithm that provides state-of-the-art performance using advanced boosting techniques and automated regularizations [13]. XGBoost combines "weak learners" and multiple classification trees in a scalable and efficient manner, providing high predictive power. Specifically, XGBoost applies additional regularisation terms, beyond a second-order Taylor approximation and gradient descent, to minimise the loss function. XGBoost is known as the fastest GB algorithm, especially for large datasets, because it can mitigate time inefficiency by evaluating feature distributions across all data points in a parallel manner, reduce overfitting by using regularisation techniques, and tune hyperparameters effectively. For all these desirable advantages — scalability, efficiency, robustness, and accuracy — XGBoost is widely used in the banking and financial industries for highly predictive credit risk modelling and mortgage default prediction.

III. ML MODEL BUILDING

The ML model development process was presented in Figure 3. The flowchart of the experimental design starts with data preprocessing, followed by sampling the data into training and testing data, feature selection, training the various classification models (see section 2.2 for detailed ML algorithms), and evaluating the model performance on both training and testing data.



[Fig.3: Diagram of Mortgage Loan Default Model Development Process]

A. Data Preprocessing

The loan-level data, extracted from Freddie Mac data and featuring a rich set of loan, borrower, and property characteristics, was appended to its loan performance. Data preprocessing and data creation were performed to construct loan statuses (i.e., whether loans were delinquent for six months or more) and other key variables for modelling. Data were cleaned, and missing values (less than 1%) were

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

addressed using mean imputation. Exploratory data analysis, correlation analysis, and feature selection were conducted to identify and select features that improve the predictions of loan defaults.

B. Sampling

The data was divided into two parts with a 75:25 split ratio. The training part, comprising 75% of the data (77,699 loans), was used to train the ML models. In contrast, the remaining 25% of the data (25,900 loans) was set aside to test the trained classification algorithms and evaluate the predictions of mortgage loan defaults.

C. Feature Selection

To support the prediction of mortgage loan defaults, feature selection was performed to identify features that can enhance the model's prediction accuracy. The twenty features selected were primary credit risk indicators in the mortgage underwriting process that can influence three main characteristics of borrowers' default behavior: credit, collateral, and capacity [14]. Among the essential variables (see Table 1), key variables include the current loan-to-value ratio (CLTV), mortgage interest rates, and borrowers' credit history (as indicated by their FICO score). Loans with high CLTV are associated with a high probability of defaults as they cause diminished incentive to make payments [15]. High mortgage rates are associated with high default risks because they intensify financial burden and instability [16]. Borrower's credit or previous behavior plays a crucial role in predicting current default behavior [17]. Other features, such as borrower and property characteristics (e.g., first-time home buyer, single-family homes, owner-occupied) and macroeconomic factors, were also included in the final dataset used for ML modelling.

D. Model Evaluation

ML models used in this study, including Logit Model, Decision Tree, Neural Network, Gradient Boosting, Random Forest, and XGBoost, were evaluated based on the following classification performance metrics: Confusion Matrix, Accuracy (ACC), Area Under the Curve (AUC), F1 scores, and Receiver Operating Characteristic (ROC). The model evaluation was performed on both training and testing data to assess whether overfitting issues occurred and to determine which model performed best.

i. Confusion Matrix

A confusion matrix is a standard statistical technique for evaluating models with binary outcomes, particularly for comparing the ML-predicted loan default status to the actual loan default status. The confusion matrix presents four categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). For example, TP refers to loans that were both predicted to default and observed to default. FP refers to loans that did not default in reality, but were incorrectly classified as defaults.

ii. Accuracy (ACC)

Accuracy is a measure of the number of correct predictions. A higher accuracy of the model is defined as a higher total number of TP and TN relative to the total number of loans assessed, indicating a higher number of defaulted and non-defaulted loans that were accurately predicted. A ML model that has a higher percentage of defaulted loans and repaid loans that were correctly classified is a model with higher prediction accuracy.

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad \ldots \quad (3)$$

iii. F1 scores

The Recall metric measures the proportion of actual loan defaults that are correctly identified among the loans that are actually defaults. Recall is also known as sensitivity or actual positive rate. The Precision metric calculates the percentage of loan defaults that are correctly predicted among the loans that are expected to default. Since evaluating Precision and Recall separately can make it difficult to compare model performance across multiple ML models during validation, the F1 score (i.e., a harmonic mean of Precision and Recall) offers a single balanced metric to rank the accuracy of various ML models. F1 scores can recognize the extreme differences in Precision and Recall when comparing the two ML models.

$$Recall = \frac{TP}{TP+FN} \quad \ldots \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad \ldots \quad (5)$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \quad \ldots \quad (6)$$

iv. Receiver Operating Characteristic (ROC)

The ROC curve is a statistical tool used to examine classification performance by ranking positive outcomes higher than negative ones across all possible thresholds. The ROC is a plot that illustrates the trade-off between TP rates and FP rates at multiple thresholds.

v. Area Under the Curve (AUC)

The Area Under the Curve (AUC) indicates the discrimination ability between positive vs. negative outcomes. AUC is a popular statistic for measuring classification accuracy. The higher the area under the ROC curve, the higher the predictive precision. An AUC of 0.5 illustrates that a model has no predictive power. Meanwhile, an AUC of 1.0 indicates that the model achieves perfect performance in classifying loans as either "good" or "bad" with 100% accuracy. Therefore, the AUC score is also a robust non-threshold accuracy metric to rank model performance, which was used in this study.

IV. RESULTS AND DISCUSSION

The study examined six ML algorithms to predict mortgage loan defaults: Logit Model, Decision Tree, Neural Network, Gradient Boosting, Random Forest, and XGBoost. Figure 4, Figure 5, and Table 2 illustrate both graphical and numerical comparisons of confusion matrices, accuracy, AUC, F1, and ROC curves across ML models.

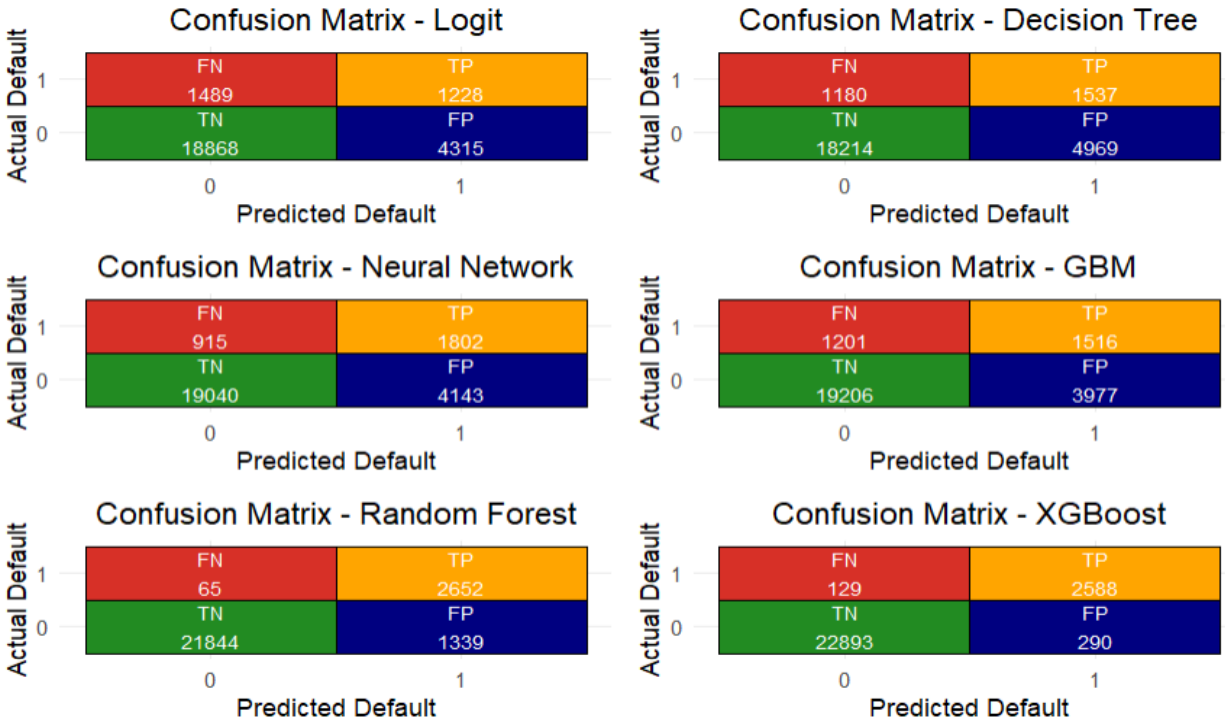
All classification algorithms demonstrate predictive ability in classifying mortgage loans as either default or non-default. However, the results showed that Random Forest and XGBoost outperformed other models, achieving superior accuracy ranging from 95% to 99%, on both training and testing data. Comparing the performance of Random

© Copyright: All rights reserved.

Forest and XGBoost, Random Forest appeared to exhibit some overfitting issues, whereas XGBoost yielded more accurate results across all metrics (accuracy, AUC, F1, and ROC). Overall, XGBoost was the top performer, achieving the highest accuracy (99%), and its performance metrics were all above 90%.

A. Results from the Confusion Matrix

Figure 4 compares confusion matrices from the six ML models in this study. 1,228 loan defaults (TP) and 18,868 non-default loans (TN), totalling 20,096 loans in the testing data. For example, the Logit Model can correctly predict 1,228 loan defaults (TP) and 18,868 non-default loans (TN), given a total of 25,900 loans in the testing data. However, among misclassifications from the Logit Model, 4,315 non-default loans were incorrectly predicted as defaults (FP) while 1,489 default loans were misclassified as non-defaults (FN). The logit model showed the worst performance among all models, illustrating the model's limitations in recognising nonlinear patterns in the data. Decision Tree, GBM, and Neural Network Models demonstrated some improved performance over the Logit Model, exhibiting higher TP and TN. However, instances of FN and FP remained relatively high. Random Forest and XGBoost Models achieved higher accuracy than the baseline Logit Model and other ML models in predicting loan defaults. Random Forest correctly predicted 2,652 loan defaults and 21,844 non-default loans, whereas XGBoost precisely predicted 2,588 loan defaults and 22,893 non-default loans. In addition, the number of FP was substantially lower for XGBoost than for Random Forest. Overall, Random Forest and XGBoost had the highest model accuracy and the lowest overall misclassification compared to other ML models on the tested data.



[Fig.4: Comparison of Confusion Matrices Across Machine Learning Models: Logistic Regression, Decision Tree, Neural Network, Gradient Boosting, Random Forest, and XGBoost on Testing Data]

B. Overall Accuracy (ACC, F1 and AUC)

Table 2 presents the comparative performance of various ML models for mortgage loan default prediction evaluated on both training and testing data. A comparison of key evaluation metrics —ACC, AUC, and F1 — demonstrated that XGBoost outperformed other ML models. The logit model played the statistical role of the traditional benchmark. While it provides a reasonable ACC (0.78), indicating some potential in classifying loan defaults, its AUC (0.64 and 0.63) and F1 (0.30) scores suggest limitations in capturing the complex patterns of default behaviours. Neural Network, Decision Tree, and Gradient Boosting exhibited some improvement in loan default predictions compared to the Logit Model. However, their accuracy scores showed only a modest improvement, reaching approximately 80%, while AUC values remained around 70%, and F1 scores were relatively low, ranging from 34% to 41%.

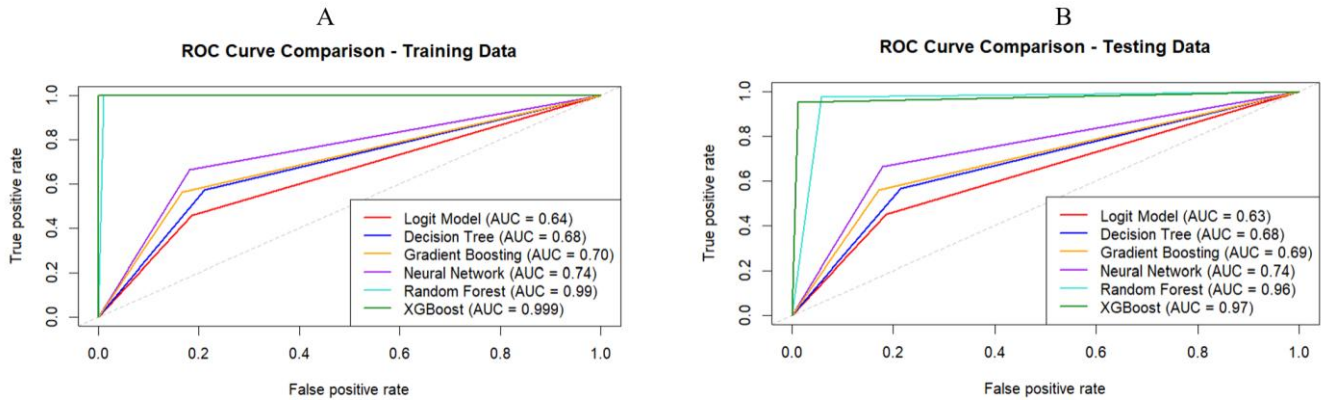
Table-II: Model Results Comparison

	Training			Testing		
Model	ACC	AUC	F1	ACC	AUC	F1
Logit	0.78	0.64	0.30	0.78	0.63	0.30
Decision Tree	0.77	0.68	0.34	0.76	0.68	0.33
Neural Network	0.80	0.74	0.41	0.81	0.74	0.42
Gradient Boosting	0.81	0.70	0.38	0.80	0.69	0.37
Random Forest	0.99	0.99	0.96	0.95	0.96	0.79
XGBoost	0.99	0.99	0.99	0.98	0.97	0.93

Both Random Forest and XGBoost exhibit the highest

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

accuracy rate (0.99) and AUC (0.99) on the training data. These results suggest that both models are the strongest candidates for predicting mortgage loan defaults. However, Random Forest's F1 scores ranked second (0.96 in training and 0.79 in testing) among all models. Meanwhile, XGBoost achieved the highest F1 scores (0.99 and 0.93) on both the training and testing datasets. The test data results highlighted that XGBoost achieved the highest precision in detecting default behaviour and maintained a superior balance between precision and recall, as evidenced by its highest F1 score in both datasets. The significant difference in F1 scores between XGBoost and Random Forest (0.93 vs. 0.79) suggests that XGBoost can harmoniously predict both TP and TN, resulting in high precision and recall (or sensitivity) rates (90% and 95%). On the other hand, Random Forest had the highest sensitivity rate in detecting default behavior (98%). However, it misclassified a greater number of non-default loans as defaults, causing a higher FP compared to XGBoost (1339 vs. 290, see Figure 4), and a significantly lower precision (66%) than XGBoost. This result also suggests that Random Forest appears to have an overfitting issue, as indicated by the significant performance gaps between the training and testing data. Overall, the XGBoost Model achieved the best performance, with the highest accuracy, AUC, and F1 on both training and testing data: ACC (0.99 and 0.98), AUC (0.99 and 0.97), and F1 (0.99 and 0.93). All performance metrics above 90% demonstrated that XGBoost had an overall excellent performance and ranked as the top performer with the highest predictive power in classifying mortgage loan defaults among all ML models in this study.



[Fig.5: ROC and AUC Comparison Across Machine Learning Models: Logistic Regression, Decision Tree, Neural Network, Gradient Boosting, Random Forest, and XGBoost. Figure 5A shows the ROC Curve Comparison on the Training Data, and Figure 5B shows the ROC Curve Comparison on the Testing Data]

C. ROC and AUC Performance

Figure 5 shows further comparison of model performance, ROC, and AUC, across various ML algorithms in this research. XGBoost and Random Forest achieved the best prediction accuracy in classifying mortgage loan defaults. Logit model has potential in classifying loan defaults, however, with the lowest predictive accuracy (AUC of 0.64 on the training data, and 0.63 on the testing data). The Neural Network, Decision Tree, and Gradient Boosting models (with an AUC of around 0.7) had slightly stronger performance than the Logit model. However, these models have significantly lower accuracy than Random Forest and XGBoost, indicating some challenges in accurately classifying mortgage loan defaults and requiring additional model refinement to capture specific data patterns. Random Forest exhibited high accuracy, with an AUC of 0.99 on the training data and 0.96 on the testing data, suggesting a significant improvement in classifying mortgage loan defaults. Last but not least, XGBoost achieved the highest AUC (0.999 on the training data and 0.97 on the testing data), demonstrating a superior performance to other ML models, based on its efficiency and accuracy (AUC and ROC).

Overall, XGBoost demonstrated superior performance compared to other ML models across multiple accuracy performance metrics (ACC, F1, AUC, and ROC). All evaluation metrics for XGBoost are not only the highest on the training data (0.99), but also the highest among all ML algorithms in the study on the testing data (ACC of 0.99, AUC of 0.97, and F1 score of 0.93). Both Random Forest and XGBoost rank the highest in terms of accuracy (ACC) on the training data because they account for various nonlinear and special data patterns in complex, high-dimensional, and large data structures, and can efficiently minimise loss or errors. It suggests that Random Forest and XGBoost are the strongest candidates for predicting mortgage loan defaults and other credit scoring models in the financial industry. However, based on the comparative analysis of all ML algorithms in this study, Random Forest seems to have some overfitting issues. In contrast, XGBoost achieved the best performance across all evaluation metrics. In summary, XGBoost demonstrated superior performance compared to other ML models and was considered a champion model in predicting mortgage loan defaults in this study.

V. CONCLUSION

The comparative analysis demonstrated that the XGBoost model outperformed other models in predicting mortgage loan defaults, using a sample of over 100,000 loans and a rich set of loan characteristics from Freddie Mac data. The XGBoost model was compared to other ML models, including Logistic Regression, Neural Network,

© Copyright: All rights reserved.

Decision Tree, Gradient Boosting, and Random Forest, by using various evaluation metrics: Accuracy, AUC, F1 scores, and ROC curves. The paper found that among all ML algorithms, XGBoost and Random Forest achieved the highest accuracy, exhibiting excellent performance. However, Random Forest appears to have an overfitting issue. XGBoost demonstrated the best performance, offering superior accuracy and precision.

The finding suggests that other ML models can offer higher accuracy than traditional logistic regression in predicting mortgage loan defaults. These ML models can detect nonlinear and special patterns in borrower credit profiles to improve model prediction accuracy. Among all ML algorithms, XGBoost demonstrated the highest predictive power and accuracy, making it well-suited for future mortgage loan default detection and classification in the financial and technology industries. This AI/ML algorithm would help identify "bad" loans promptly, minimise potential loss, and assist professionals in large banks and the financial sector in making more effective and accurate decisions in credit risk assessment.

In conclusion, this study contributes an innovative FinTech solution, demonstrating that XGBoost not only achieves exceptional performance in predicting mortgage loan defaults but also offers broader potential as a superior classification algorithm for diverse fields.

ACKNOWLEDGMENT

I would like to thank my AP Statistics and Computer Science teachers, as well as the online instructors from Harvard and Stanford, for their guidance in Data Science, Machine Learning, and AI using R and Python. They helped me discover my strong interests in data science, finance, statistics, AI/ML models, and innovative technology. I would like to extend my sincere thanks to Professor Witold Henisz and Professor Anne Jamison from the University of Pennsylvania for inspiring my research interests in finance and the FinTech industry.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/Competing Interests:** Based on my understanding, this article does not have any conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it was conducted without any external influence.
- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The data and materials supporting the findings of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications*. Society for Industrial and Applied Mathematics. DOI: https://doi.org/10.1137/1.9781611974560
2. Adelino, M., Schoar, A., & Severino, F. (2018). The Role of Housing and Mortgage Markets in the Financial Crisis. *Annual Review of Financial Economics*, *10*(1), 25–41. DOI: https://doi.org/10.1146/annurev-financial-110217-023036
3. Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine Learning-Driven Credit Risk: A Systematic Review. *Neural Computing and Applications*, *34*(17), 14327–14339. DOI: https://doi.org/10.1007/s00521-022-07472-2
4. Freddie Mac. (2025). *Single Family Loan-Level Dataset*. Www.freddiemac.com. https://www.freddiemac.com/research/datasets/sf-loanlevel-dataset
5. Chen, D., Ye, J., & Ye, W. (2023). Interpretable selective learning in credit risk. *Research in International Business and Finance*, *65*(0275-5319), 101940–101940. DOI: https://doi.org/10.1016/j.ribaf.2023.101940
6. Park, Y. S., Lek, S., & Erik, J. S. (2016). Chapter 7: Artificial Neural Networks - Multilayer Perceptron for Ecological Modelling. In *Developments in Environmental Modelling* (Vol. 28, pp. 123–140). Elsevier. DOI: https://doi.org/10.1016/B978-0-444-63623-2.00007-4
7. Duan, J. (2019). Financial system modelling using deep neural networks (DNNs) for practical risk assessment and prediction. *Journal of the Franklin Institute*, *356*(8), 4716–4731. DOI: https://doi.org/10.1016/j.jfranklin.2019.01.046
8. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. DOI: https://doi.org/10.1201/9781315139470
9. Breiman, L., Cutler, A., Liaw, A., Wiener, M., & Liaw, M. A. (2018). Package 'randomforest'. University of California, Berkeley: Berkeley, CA, USA, 81, 1-29. DOI: https://doi.org/10.32614/CRAN.package.randomForest
10. Dixon, M. F., Halperin, I., & Bilokon, P. A. (2020). *Machine Learning in Finance: From Theory to Practice*. Springer. DOI: https://doi.org/10.1007/978-3-030-41068-1
11. Levantesi, S., & Piscopo, G. (2020). The Importance of Economic Variables on the London Real Estate Market: A Random Forest Approach. *Risks*, *8*(4), 112. DOI: https://doi.org/10.3390/risks8040112
12. Lu, H., Karimireddy, S. P., Ponomareva, N., & Mirrokni, V. (2020). Accelerating gradient boosting machines. In International Conference on Artificial Intelligence and Statistics (pp. 516-526). PMLR, https://proceedings.mlr.press/v108/lu20a.html
13. Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'16*, *1*(1), 785–794. DOI: https://doi.org/10.1145/2939672.2939785
14. Goel, A., & Rastogi, S. (2021). Understanding the Impact of Borrowers' Behavioural and Psychological Traits on Credit Default: A Review and Conceptual Model. *Review of Behavioral Finance*, *15*(2). DOI: https://doi.org/10.1108/rbf-03-2021-0051
15. Otero González, L., Durán Santomil, P., Lado Sestayo, R., & Vivel Búa, M. (2016). The impact of loan-to-value on the default rate of residential mortgage-backed securities. *The Journal of Credit Risk*, *12*(3). DOI: https://doi.org/10.21314/jcr.2016.210
16. Ahmad, F., & Shehzad, C. T. (2024). The Role of the Interest Rate Environment in Mortgage Pricing. *International Review of Economics & Finance*, *89*, 225–245. DOI: https://doi.org/10.1016/j.iref.2023.07.102
17. Li, Y., & Shi, Y. (2024). Credit Evaluation System Based on FICO. *Applied and Computational Engineering*, *96*(1), 48–55. DOI: https://doi.org/10.54254/2755-2721/2024.17854

AUTHOR'S PROFILE



Chi Ton Cong Nguyen is a senior at James Madison High School, USA. His research areas include AI, Machine Learning, Computer Science, Big Data, Macroeconomics, and Finance. His research interest is to design innovative AI and Machine Learning Models with high predictive power in the FinTech industry. Proficient in R, Python, and other computer software, he brings a solid command of Statistics, Data Modelling, and Computer Programming to optimise large-scale data management for financial research and data insights. His experience includes developing AI/ML predictive algorithms to enhance financial decision-making processes,

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

conducting financial forecasting and credit risk assessments, and publishing a series of research papers on AI/ML models for practical use in the financial industry. He would like to be in rigorous academic environments focused on innovation, research, and impact.



Dr. Victoria is an independent researcher with a Ph.D. from Michigan State University, USA, and over 20 years of experience in teaching and research at several prominent U.S. institutions and universities. Her research spans a broad range of fields, including Artificial Intelligence, Machine Learning, Data Science, Real Estate, House Price Econometric Models, and Financial and Technological Innovation. She is particularly well-regarded for her in-depth expertise in Statistics, Time Series and Panel Data Econometrics, as well as AI/ML models. Her guidance has proven to be top-notch for students and professionals, particularly in financial and real estate research. She has been an extraordinary mentor who can make a lasting impact on academics and applied research for anyone pursuing advanced work in Statistics, Econometrics, Economics, Finance, or Predictive Modelling.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.