Default Prediction in the Finance Industry Based on Ensemble Learning: Combining Machine Learning and Deep Learning

Le Hoanh-Su, Le Quang Chan Phong, Truong Cong Vinh, Ho Mai Minh Nhat
University of Economics and Law, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
Jong-Hwa Lee
Dong-Eui University, Busan City, South Korea

Abstract

Background: Financial institutions face significant challenges in predicting loan defaults, which directly impact the non-performing loan (NPL) rate. Incorrect predictions can lead to misinformed decisions and substantial financial losses. **Objectives:** This study aims to enhance default prediction by employing advanced ensemble learning techniques in machine learning and deep learning. **Methods/Approach:** Instead of relying on transformation, fine-tuning, or single algorithm models, this research focuses on combining multiple models using voting and stacking techniques, particularly highlighting a stacking model combining Light Gradient Boosting Machine (LGBM) and Artificial Neural Networks (ANN). **Results:** The ensemble learning methods, especially the LGBM-LSTM and XGB-LSTM stacking models, showed higher precision in identifying borrowers who defaulted, while the LGBM-LSTM and XGB-LSTM voting models excelled in recall and achieved an F1-score 0.1% higher. Both the stacking and voting models attained AUC values close to 90%, indicating strong overall classification performance. **Conclusions:** The findings not only contribute to the fields of lending and peer-to-peer financial operations but also offer crucial insights that aid financial organizations in making well-informed decisions regarding loan processing and management.

Keywords: default prediction; risk assessment; machine learning; deep learning; ensemble learning; online lending

JEL classification: G17, G21, G30, O32
Paper type: Research article

Received: 5 May 2024
Accepted: 11 Oct 2024

Citation: Le, H.S., Le, Q.C.P., Truong, C.V., Ho, M.M.N., Lee, J.H. (2025). Default Prediction in the Finance Industry Based on Ensemble Learning: Combining Machine Learning and Deep Learning. Business Systems Research, 16(1), 198-218.
DOI: https://doi.org/10.2478/bsrj-2025-0010
Corresponding author: Jong-Hwa Lee, Email: jhlee6050@deu.ac.kr

Introduction

In the field of finance and banking, the importance of lending is matched by the importance of risk management. There are too many loan applications every day, causing difficulties for employees to process, and wrong decisions can lead to huge losses to banks or other financial organizations (Gupta et al., 2020). Therefore, evaluating the repayment ability of customers is currently a top priority for global financial institutions (Kim et al., 2020).

Amid the current economic challenges, the economies of many countries are declining, leading to a decrease in the demand for loans from individuals and organizations. A notable example is the credit balance in Vietnam, which suffered a decreasing trend in the whole country (Chi Tin, 2023). The decision of banks and credit organizations to focus on developing various lending sectors requires careful consideration to attract customers and promote credit growth. However, this can come with risks of bad debts and difficult recoveries due to various reasons, such as insufficient understanding or incomplete information about customers and subjective judgments in the loan application assessment process. On the other hand, decisions based on incomplete information may cause businesses to miss out on potential customers, those with the ability to borrow and repay loans. Therefore, to ensure the quality of assessments, it is essential to evaluate loan applications from multiple perspectives, considering customer information and credit history, and minimizing human-based evaluations.

The limits of standalone models such as K-Nearest Neighbors (KNN) and Logistic Regression (LR) are becoming more evident in the developing field of credit risk assessment, particularly when dealing with complicated and non-linear data interactions (Chen et al., 2023; Ali et al., 2023). Even while these conventional models are appreciated for their simplicity and readability, as dataset volumes increase, they frequently fail to provide the prediction accuracy required (Acito, 2023). By utilizing the advantages of several base models, ensemble techniques like voting and stacking, on the other hand, provide more reliable and accurate predictions (Dong et al., 2020; Rincy & Gupta, 2020). For instance, Kumari et al. (2021) proved the efficacy of a Soft Voting classifier that combines Random Forest (RF), Naive Bayes (NB), and LR. (Kim et al., 2020) also found that using deep learning models greatly enhanced loan repayment predictions in peer-to-peer lending. Stacking and voting approaches can result in significant performance gains, as demonstrated by (Uddin, et al., 2023) outstanding accuracy rate of over 93% in their bank loan approval system. By adopting these approaches, financial institutions can better balance interpretability with predictive power, ultimately leading to more effective risk management in peer-to-peer lending platforms (Uddin et al., 2023; Emmanuel et al., 2024).

This study aims to predict the loan default of online loan customers by applying staking and voting techniques in lending activities. The results of the study show that the stacking model provides superior accuracy. The study consists of five sections: beginning with an introduction to lending concepts and background in sections 1 and 2, the research will delve into the basis and application of machine learning and deep learning techniques that are applied in this research in section 3. The study will conclude with an analysis of experimental results derived from the implemented models in sections 4 and 5. This outcome provides a significant contribution to and propels advancements in the realm of online lending, offering tangible value and fostering growth in the fields of credit and finance.

Literature Review

Lending

According to Law on Credit Institutions of Vietnam Governance (2010), lending is a process in which an individual, organization, or agency (the lender) provides a sum of money to another individual, organization, or agency (the borrower) with the understanding that the borrower will repay the principal amount along with interest after a certain period.

Lending can take place in various forms, including bank lending, lending through financial companies, or peer-to-peer lending. Each form of lending has its characteristics, regulations, and risks.

Along with the development of information technology, peer-to-peer lending activities are also rapidly expanding. This is a financial service model using Fintech technology to create a direct connection between borrowers and investors without the intervention of intermediary financial institutions such as banks (Basha et al., 2021).

A loan default will happen anytime a borrower fails to meet the payment commitments or other conditions stipulated in the agreement of a loan contract, which typically leads to being unable to repay either some or all of what is owed on the principle and interest (Satpute, et al., 2022; Koç & Sevgili, 2020). According to Adedapo (2007), loan default is defined as the inability of a borrower to fulfill his or her loan obligation as and when due. Zhou (2023) highlighted that loan default prediction forecasts the probability of default based on the information already available about the loan applicant and determines whether to release the loan. Therefore, default prediction is critical for financial institutions and investors.

Predicting loan default capacity is crucial in the financial sector for assessing risk and making lending decisions. Accurate forecasting helps minimize risk, enhance business efficiency, and identify potential customers. However, this process faces challenges such as collecting and processing complex data, requiring deep financial and data analysis knowledge. Recently, the surge in machine learning (Jayaram et al., 2024) and big data analytics (Abhiram et al., 2023) has opened new opportunities for predicting default risk, especially in Peer-to-Peer (P2P) lending. Despite these advancements, risks related to data quality and economic fluctuations remain significant challenges. The effectiveness of prediction models heavily depends on the quality of input data; issues like missing data and complex preprocessing can significantly affect prediction accuracy (Fang & Ji, 2024). Additionally, the continuously evolving financial environment with rapid economic changes can alter borrower behavior, complicating the prediction process (Dhruv et al., 2023).

Defaults can arise from various factors. Studies have shown that a combination of personal characteristics, loan details, and macroeconomic conditions influences the likelihood of default. Qi (2023) identified several critical factors impacting default probability, including borrower attributes (such as income level and credit history), loan characteristics (like loan amount and interest rate), and macroeconomic factors (such as unemployment rate and inflation). Analyzing these variables suggests that financial institutions can enhance risk assessment models by integrating these factors into credit scoring systems.

Standalone Model Approaches

Traditional Methods

LR is a widely applied statistical method in finance, especially for predicting credit risk. The model predicts the probability of a binary event, such as whether a customer is capable of repaying a loan (1) or not (0) (Chen et al., 2023). With its simplicity and interpretability, LR is often the preferred choice in many studies, particularly when the dataset is small or when the independent variables are not highly correlated (Chen et al., 2023; Schonlau, 2023). However, LR's predictive power may be limited in complex or non-linear problems, which often require more advanced models.

KNN is a supervised machine learning algorithm that is easy to understand and implement. The core principle of KNN is based on proximity, meaning similar data points are located near each other in feature space (Ali et al., 2023; Uddin et al., 2022). When classifying a new data point, KNN identifies the k nearest neighbors in the training set and assigns a label based on the majority vote of these neighbors. While KNN is suitable for both classification and regression tasks, it can become inefficient with large datasets due to the high computational cost and memory requirements. The choice of k directly affects model performance, necessitating careful tuning (Acito, 2023).

Jayaram, Balachandar, and Kumar (2024) utilized the LR algorithm in machine learning to develop a credit scoring model for Peer-to-Peer (P2P) lending platforms. The study successfully predicted default probability, improving the differentiation of high and low-risk borrowers. The results demonstrate significant improvement in credit risk prediction compared to traditional methods, contributing to more effective and reliable risk assessment tools for P2P lending.

Advanced Methods

Extreme Gradient Boosting (XGBoost) is a variant of the Gradient Boosting algorithm, proposed by Chen and Guestrin (2016). XGBoost is known for its optimized computational efficiency and high prediction accuracy. The algorithm builds a sequence of models, where each subsequent model corrects the errors of the previous one, leading to improved prediction quality (Brownlee, 2016; Hakkal & Lahcen, 2024). XGBoost can handle large, complex datasets at high speed, and its flexibility in parameter tuning makes it a popular choice in many machine learning competitions. However, improper tuning of XGBoost can lead to overfitting, resulting in poor performance on unseen data.

Light Gradient Boosting Machine (LGBM) is a boosting algorithm developed and published by Microsoft in the paper by Ke et al. (2017). A key feature of LGBM is its leaf-wise growth strategy, which optimizes tree building better than traditional level-wise tree structures (XGBoost), thus speeding up training and reducing errors (Ke et al., 2017; Machado et al., 2019). Thanks to its memory efficiency and high performance, LGBM is often applied to large datasets with numerous features. To achieve optimal performance, users must carefully adjust model parameters to balance optimization with the model's ability to generalize to new data.

The structure and functioning of the human nervous system inspire Artificial Neural Networks (ANN). They are designed to model non-linear relationships in data (Thorat et al., 2022; Alaloul & Qureshi, 2020). ANN consists of multiple layers of artificial neurons, where each layer learns different features from the data, allowing the model to capture complex patterns. ANN is particularly powerful in handling unstructured and complex data, such as images and text, thanks to its ability to learn and optimize weights automatically (Pandey & Pandey, 2022). However, ANN demands significant computational resources and can be challenging to implement effectively without careful tuning (Sain & Kumar, 2022).

Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Networks (RNN) introduced by Hochreiter & Schmidhuber (1997). LSTM addresses issues in processing sequential data and learning long-term dependencies by using gates to regulate information storage and forgetting. This allows LSTM to retain important information during training, making it an ideal choice for time series analysis, machine translation, and speech recognition (Wang et al., 2018; Yadav, et al., 2023). Despite its advantages, training LSTM models requires considerable time and computational resources, particularly with large datasets.

Wang et al., (2024) explored the effectiveness of the XGBoost algorithm in predicting credit risk. The research compared XGBoost with other machine learning methods like KNN, RF, and LR, concluding that XGBoost excels in identifying credit risks based on borrower characteristics. This study integrates into the credit risk assessment field by highlighting XGBoost's capability to handle complex factors affecting repayment ability. While achieving higher accuracy, the study also points out areas for improvement in XGBoost, indicating potential for further development in optimizing credit scoring systems using advanced machine learning algorithms.

Fan (2023) introduced a personal loan default prediction platform using the LGBM algorithm, emphasizing its superior performance compared to models like RF. The research shows that LGBM achieves higher accuracy, precision, and discrimination in default prediction. This advancement is significant for personal loan risk assessment, aiding financial institutions in making more informed credit decisions. Notably, the study extends beyond model development to implement a real-time prediction system using Python and Flask. This system allows for quick input of customer data and prediction of default probability, enhancing credit risk management processes.

Ensemble method

Ensemble Learning is a machine learning approach where multiple models, known as base learners, are trained to solve the same problem. The final result is determined by aggregating predictions from these models, improving predictive performance compared to using a single model alone (Dong et al., 2020; Rincy & Gupta, 2020). The fundamental principle of Ensemble Learning is that combining different models can leverage the strengths of each base learner, producing a more accurate overall model.

Stacking is an advanced technique in Ensemble Learning, introduced by Wolpert (1992). Unlike other Ensemble methods, Stacking uses the predictions from multiple models (base learners) as input for a higher-level model known as a meta-learner. Rather than simply averaging or voting on predictions, the meta-learner is trained to "learn" how best to combine the information from these predictions to produce the final optimal outcome (Li et al., 2020; Kalule, Abderrahmane et al., 2023).

By intelligently combining the base models, Stacking can reduce both bias and variance, depending on the construction and choice of base learners (Li et al., 2019). Stacking is especially effective when base models have different structures and prediction methods, creating diversity within the ensemble.

Voting is a popular Ensemble technique where the final result is determined by aggregating predictions from multiple base models through a "voting" process (Kabari & Onwuka, 2019). Unlike Stacking, models in Voting operate independently, without learning from each other. Instead, each model's prediction is considered, and the final decision is made based on either majority voting (hard voting) or probability averaging (soft voting).

In Hard Voting, each model casts a "vote" for a label, and the label with the most votes is selected as the final prediction. This technique is suitable for classification problems with multiple classes and when base models have relatively similar accuracy levels. Soft Voting, on the other hand, aggregates the probability predictions from base models. The label with the highest cumulative probability is selected as the final prediction. This method is more effective when base models can output reliable probability estimates, even if their accuracies differ (Fauzi & Yuniarti, 2018).

Emmanuel et al. (2024) presented an advanced credit risk prediction system using a stacking model combined with feature selection based on Information Gain (IG). By addressing class imbalance in credit risk datasets, the authors demonstrated that the stacking model with Gradient Boosting, RF, and XGBoost significantly improves risk classification. With AUC scores up to 0.944 across various datasets, this approach shows superiority over individual models. Integrating advanced machine learning methods like stacking enhances the accuracy and effectiveness of credit risk assessment systems, supporting financial institutions in more effective and sustainable risk management.

The paper by Uddin et al. (2023) investigates the use of machine learning for bank loan approval prediction. Significantly, rather than depending on a single machine learning model, the authors utilized the Ensemble Voting technique, which combines the predictions from the three most effective models: RF, Extra Trees (ET), and KNN. The rationale behind selecting the Ensemble Voting technique lies in its capacity to harness the strengths of each model while simultaneously diminishing their weaknesses. Empirical results demonstrate that the Ensemble model attained an accuracy of 87.26%, surpassing the performance of any individual component model and exceeding the results reported in prior studies cited in the paper. This research substantiates the efficacy of Ensemble Learning in addressing the task of loan approval prediction, thus presenting a novel avenue for financial institutions to enhance prediction accuracy and mitigate risks associated with lending decisions.

Comparison of Standalone versus Ensemble Methods

To predict default lending in peer-to-peer (P2P) lending as well as banking platforms, prediction methods have been developed from traditional statistical methods, such as LR, KNN, to advanced machine learning and deep learning techniques. LR and KNN are favored for their simplicity, interpretability, and high explainability, particularly when data has a linear structure or simple relationships between variables (Chen et al., 2023; Ali et al., 2023). However, these models become limited when faced with data containing complex non-linear relationships or when the dataset size increases, leading to reduced scalability and prediction accuracy (Acito, 2023). Hence, modern techniques have emerged, such as XGBoost, LGBM, ANN, and LSTM. These methodologies approach work extremely well in the handling of nonlinear relationships, automatic feature extraction via datasets, and improving computational efficiency on complex problems with huge datasets. However, such advanced models usually require measurable computational resources and suffer from data overfitting problems if not tuned well (Fan, 2023; Wang et al., 2024).

The key distinction between the two groups lies in accuracy and generalizability. While traditional statistical models generally offer lower accuracy in handling complex, non-linear data, they remain easier to interpret and explain. In contrast, modern models give better accuracy but are not interpretable and have higher training times.

To overcome these limitations, Ensemble Learning combines multiple models to improve both prediction accuracy and reliability while optimizing the explainability and predictive power tradeoff. Techniques like stacking and voting harness the strength of multiple base models, minimizing bias and variance, and resulting in more robust and stable systems. Ensemble Learning not only leverages the strengths of both traditional and modern methods but also provides a balanced solution that enhances credit risk prediction capabilities in P2P lending systems (Uddin, et al., 2023; Emmanuel et al., 2024). In summary, Ensemble Learning offers a comprehensive approach to addressing the complexities and performance challenges in credit prediction, ultimately supporting more effective financial decision-making.

Methodology

Experimental Procedure

This experimental procedure begins with data collection, which meticulously gathers valuable information from reliable sources. Subsequently, we proceed to prepare the data. This step is crucial as it lays the foundation for effective model training as presented in Figure 1.

Next, model training becomes a critical step, where we focus on utilizing Ensemble Learning methods such as the Stacking Model and Voting Model. Specifically, with the Stacking Model, we employ a series of meta-learner models like LG, KNN, XGBoost, LGBM to learn the best way to combine predictions from various base models such as ANN and LSTM. Similarly, with the Voting Model, we also use this combination approach to make predictions from different models by using majority voting or weighted voting methods. After training the model, we continue with the step of 'Predicting Loan Default according to customer attributes" by using algorithms to predict whether a loan will default based on the attributes of the customer. Finally, evaluation and conclusion are the last steps, where we interpret the results, assess the model's performance, and derive profound insights from empirical evidence and model explanations.

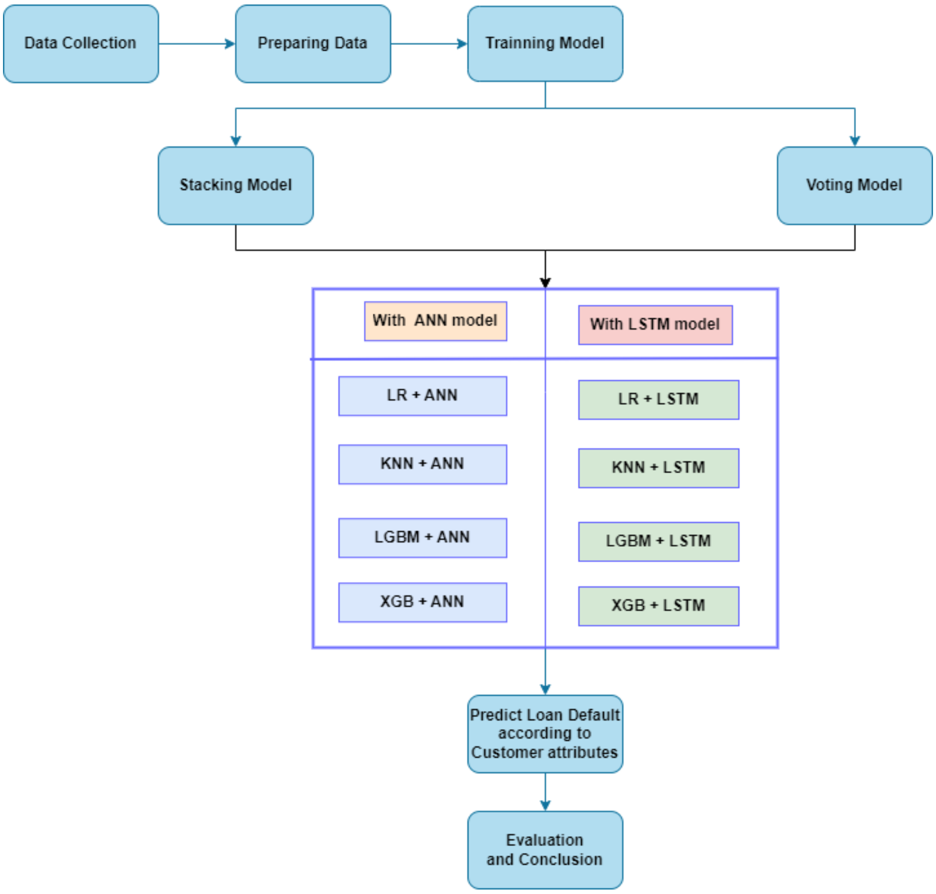
Model effectiveness evaluation method

An important step after building a model is evaluating its performance and quality using evaluation methods. In this study, the authors used metrics such as Accuracy, Precision, Recall, F1-score, and AUC ROC to compare and select the most suitable model for predicting customer loan repayment within the dataset.

In the context of the application data in the lending field in this study, the authors classify customers who cannot repay the loan as Positive and those who can repay as Negative. Each prediction can fall into one of four outcomes, based on how it matches the actual value:

- True Positive (TP): Customers who are predicted not to have the ability to repay their debts (Positive) and actually do not repay.
- True Negative (TN): Customers who are predicted to have the ability to repay their debts (Negative) and actually do so.
- False Positive (FP - Type 1 error): Customers who are predicted not to have the ability to repay their debts (Positive) but actually do repay.
- False Negative (FN - Type 2 error): Customers who are predicted to have the ability to repay their debts (Negative) but actually do not repay.

Figure 1
Experimental Procedure



Source: Author's work

Precision reflects the reliability of a machine learning model in predicting samples as Positive. This means that among the samples that the model predicts as Positive, what percentage is actually Positive? The higher the Precision, the more confident we can be that the samples predicted by the model as Positive are indeed Positive. The formula is expressed in the study by Sokolova et al. (2006):

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (or Sensitivity) evaluates the completeness of a machine learning model in predicting samples as Positive. It tells us how much of the actual Positive samples the model has 'recalled'. The higher the Recall, the better the model is at finding actual Positive samples in the data. The formula is expressed in the study by Sokolova et al. (2006):

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score is a model evaluation metric based on the balance between Precision and Recall. In some cases, optimizing Precision can reduce Recall and vice versa. F1-score is a way to measure this balance. In other words, F1-Score helps us evaluate how well the model achieves a balance between Precision and Recall. The formula is expressed in the study by Sokolova et al. (2006):

$$F1\_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The ROC Curve is a visual graph illustrating the classification ability of a model. It shows how the model makes decisions based on different confidence levels. The ROC curve consists of two main factors: True Positive Rate (TPR), which measures the frequency of the model accurately predicting positive cases; and False Positive Rate (FPR), which measures the frequency of the model incorrectly predicting negative cases as positive. By analyzing this graph, we can evaluate the effectiveness of the model and determine the appropriate threshold to achieve the best balance between correct and incorrect predictions. The formulas for calculating TPR and FPR are given in the research work by Abdelmoula (2015):

$$TPR = \frac{TP}{TP + FN} = Recall \quad (5)$$

$$FPR = \frac{FP}{FP+TN} = 1 - Specificity \quad (6)$$

AUC (Area Under Curve), in the study by Abdelmoula (2015), is recognized as a general metric to evaluate the performance of a classification model at different classification thresholds. When AUC approaches 1, it indicates that the classification model is operating very effectively. In the case where AUC equals 0.5, it indicates that the model cannot distinguish between the Positive and Negative classes. Even if the AUC is less than 0.5, it means that the model is predicting in reverse, that is, predicting the Positive class as Negative and vice versa.

Data Collection

Based on the previous research, lending activities can be conducted through various channels such as banks, financial companies, or peer-to-peer lending platforms. For this particular study, we utilized the Lending Club dataset, referenced by George (2021) on the Kaggle platform, to predict borrowers' loan repayment capabilities. Lending Club is a peer-to-peer lending platform that connects borrowers with investors. The original dataset consisted of millions of loan records, each representing a loan issued to an individual borrower, collected from LendingClub.com between 2007 and 2016. However, to ensure the relevance and accuracy of their analysis, we focused on the two most recent years of the dataset. Therefore, the sampling frame includes only loans issued in 2015 and 2016. This filtered dataset includes 122,352 individual loan records, each with 27 attributes per loan, covering details such as loan amount, interest rate, term, borrower characteristics, and loan status as presented in Table 1.

Table 1
Attributes in Loan data

No	Attribute	Attribute Description	Example values	Non-Null Count	Data type
0	loan_amnt	Amount of Loan	1000, 8000	122,352	int64
1	term	Loan period in month	36 mons, 60 mons	122,352	object
2	int_rate	Loan interest	11.44, 17.27	122,352	float64
3	installment	The average monthly repayment amount if the loan is activated	329.48, 265.68	122,352	float64
4	grade	Group of customers according to Lending Club	B, A	122,352	object
5	sub_grade	Sub-group of customers according to Lending Club	B4, A2	122,352	object

6	emp_title	Occupation	Marketing, Credit analyst, Pilot	115,655	object
7	emp_length	Year of occupation	10+ years, <1 year, 3 years	115,648	object
8	home_ownership	Homeownership status	RENT, OWN, MORTGAGE	122,352	object
9	annual_inc	Annual income	117000, 46000	122,352	float64
10	verification_status	Verification of income status	Not Verified, Verified	122,352	object
11	issue_d	Loan issue date	Jan-15, Oct-14	122,352	object
12	purpose	Loan purpose	Vacation, credit_card	122,352	object
13	title	Title for loan purposes given by customers	Vacation, Credit card Refinance	122,352	object
14	dti	Debt-to-income ratio	26.24, 22.05	120,250	object
15	earliest_cr_line	First date, the customer opens a credit line	Jun-90, Aug-07	121,232	float64
16	open_acc	Number of opened accounts	16, 8	121,232	object
17	pub_rec	Public credit recognition	0, 1	122,352	int64
18	revol_bal	The related credit account balance of the customer	36369, 20131	122,352	int64
19	revol_util	The validity period of the credit line	41.8, 100.6	122,352	int64
20	total_acc	Total credit account	25, 40	121,232	float64
21	initial_list_status	The initial recognition status	W, F	122,352	int64
22	application_type	Application type	INDIVIDUAL, JOINT, DIRECT_PAY	122,352	object
23	mort_acc	Total mortgage account	0, 3, 4	122,352	object
24	pub_rec_bankruptcies	Public recognition of bankruptcies	0, 1	122,352	float64
25	loan_status	Loan status	Fully Paid, Charged Off	122,352	float64

Source: Author's work

Table 2

Descriptive statistics are applied to the Loan data

Variable	count	mean	std	min	25%	50%	75%	max
loan_amnt	122352	14888,98	8811,67	1000	8000	13000	20000	40000
int_rate	122352	13,43	4,72	5,32	9,99	12,99	16,55	30,99
installment	122352	445,94	261,57	30,12	255,04	383,46	593,57	1533,81
annual_inc	122352	77845,31	72187,8	0	48000	65000	93000	8706582
dti	122352	18,76	30,24	0	12,05	18,08	24,8	9999
open_acc	122352	11,95	5,62	1	8	11	15	90
pub_rec	122352	0,24	0,66	0	0	0	0	86
revol_bal	122352	16517,11	24055,45	0	5655	10836	19795,5	1023940
revol_util	122283	50,1	24,86	0	31,2	49,8	69,1	152,5
total_acc	122352	26,45	12,44	2	17	25	33	151
mort_acc	122352	1,79	2,05	0	0	1	3	34
pub_rec_bankruptcies	122352	0,16	0,41	0	0	0	0	8

Source: Author's work

The target demographic for the Lending Club loan dataset comprises borrowers who applied for and obtained loans via the Lending Club platform. The descriptive statistics of the observed units and the outline of the sampling are presented in Table

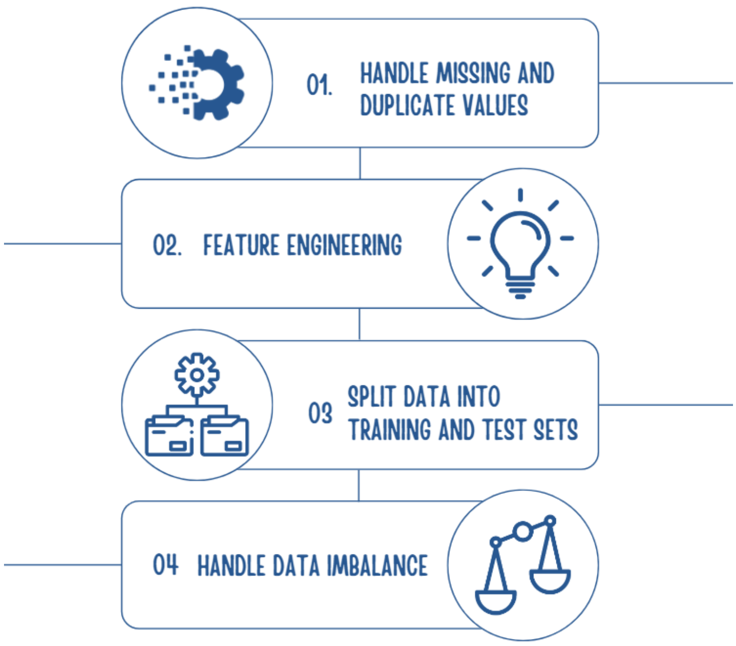
2. These borrowers typically seek loans for various purposes, such as establishing a small business or taking a trip. The dataset encompasses a diverse group of individuals with different income levels, work histories, credit ratings, and financial backgrounds. By capturing precise loan features and borrower characteristics, the dataset is beneficial for credit risk analysis and predicting repayment behavior. This group of borrowers is representative of the larger customer base that opts for peer-to-peer lending instead of traditional banking and financial services.

Data Preparation
Building predictive models requires thorough data preprocessing to ensure the data is ready for analysis and compatible with various algorithms. As shown in Figure 2, the main steps begin with handling missing and duplicate values using dropna() and drop_duplicates() to clean the data.

After that, feature engineering is performed, such as extracting "zipcode" from the "address" column to create more meaningful features. LabelEncoder() is used to convert categorical variables into a numerical format, making the data compatible with machine learning algorithms. Before applying feature selection, MinMaxScaler() is used to normalize the data, ensuring that all features are on the same scale. Then, reduce the number of input variables in the model by retaining only the most relevant features, improving model performance, and reducing complexity. Table 3 shows the features retained and included in the model.

After feature engineering, the dataset is split into training and testing sets in an 80:20 ratio to ensure robust model evaluation. Finally, due to the imbalance in the target variable (95,191 loans are labeled as "Fully Paid" (77.8%) and 27,161 loans are "Charged Off" (22.2%)), the Synthetic Minority Oversampling Technique (SMOTE) is applied to oversample the minority class and balance the dataset. This step ensures that the model does not disproportionately favor the majority class.

Figure 2
Data Preparation



Source: Author’s work

Table 3
Features retained and included in the model

No	Attribute	Attribute Description
0	loan_amnt	Amount of Loan
1	term	Loan period in month
2	int_rate	Loan interest
3	installment	The average monthly repayment amount if the loan is activated
4	sub_grade	Sub-group of customers according to Lending Club
5	home_ownership	Homeownership status
6	annual_inc	Annual income
7	verification_status	Verification of income status
8	purpose	Loan purpose
9	dti	Debt-to-income ratio
10	open_acc	Number of opened accounts
11	revol_bal	The related credit account balance of the customer
12	total_acc	Total credit account
13	initial_list_status	The initial recognition status
14	application_type	Application type
15	Zip code	Zipcode of Address
16	loan_status	Loan status

Source: Author's work

Results

Experimental Results of the Stacking Model

To optimize the process of predicting default probability, a range of advanced machine learning algorithms was applied, including LR, KNN, XGBoost, and LGBM. Additionally, the authors have leveraged the power of deep learning through the deployment of LSTM and ANN models. To enhance accuracy and generalization ability, these models have been combined through stacking methods. Below is a detailed comparison table illustrating how each algorithm combination affects the overall model performance.

Experimental Results of the Stacking ANN Model

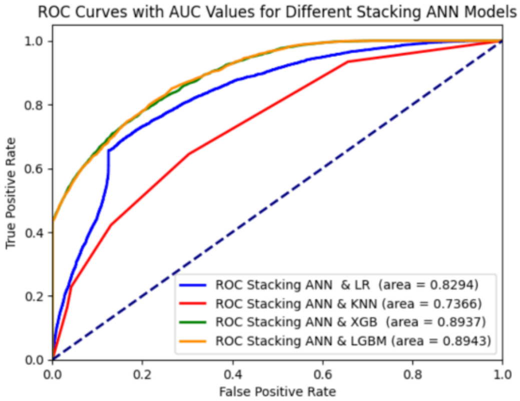
As illustrated in Table 4 and Figure 3, the Stacking LGBM + ANN and XGBoost + ANN models have high accuracy (around 77.98% and 77.07%, respectively). This indicates that the combination of boosting algorithms with ANN can significantly improve predictive capabilities compared to individual models. Notably, both models also achieve high scores in Precision and F1-Score, demonstrating their ability to detect loan defaults accurately and balance precision and recall in prediction.

Table 4
Experimental Results of Various Stacking ANN Models

Stacking ANN	LR + ANN	KNN + ANN	LGBM + ANN	XGBoost + ANN
Accuracy	82.60%	76.88%	86.81%	86.90%
Precision	75.14%	66.19%	85.33%	85.65%
Recall	76.59%	64.60%	74.48%	74.51%
F1-Score	75.81%	65.27%	77.98%	77.07%
ROC	82.49%	73.66%	89.43%	89.37%

Source: Author's work

Figure 3
The results of the ROC Curve for the Stacking ANN model



Source: Author's work

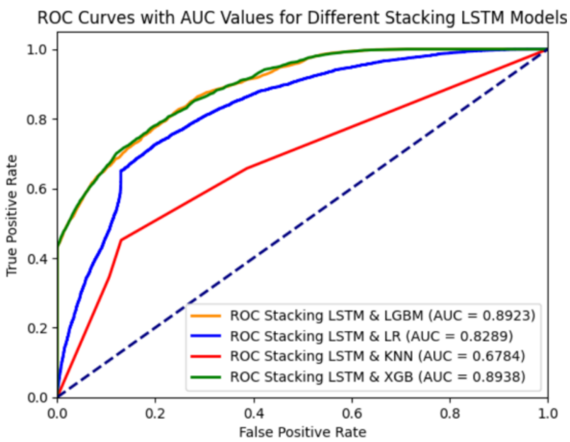
Experimental Results of the Stacking LSTM Model
As illustrated in Table 5 and Figure 4, the results from using LSTM, a deep recurrent neural network, in the stacking model demonstrate the clear superiority of boosting algorithms like XGBoost, with high F1-Score and ROC scores, over 78% and 89%, respectively. LGBM's results are also positive. Meanwhile, models combining LSTM with LR and KNN yield less impressive results, reflecting the difference in the capabilities of the underlying algorithms when applied to the same advanced network architecture like LSTM.

Table 5
Experimental Results of Various Stacking LSTM Models

Stacking LSTM	LR + LSTM	KNN + LSTM	LGBM + LSTM	XGBoost + LSTM
Accuracy	82.06%	77.53%	86,89%	86,88%
Precision	74.44%	67.35%	85,95%	85,58%
Recall	76.03%	66.05%	74,25%	74,50%
F1-Score	75.16%	66.62%	77,91%	78,04%
ROC	82.89%	67.84%	89,23%	89,38%

Source: Author's work

Figure 4
The results of the ROC Curve for the Stacking LSTM model



Source: Author's work

Experimental Results of the Voting Model

In their experimental Voting model, the author team only implemented Soft Voting. The reason is that Soft Voting calculates the average probability of predictions from the sub-models instead of choosing a class based on the majority of votes, like Hard Voting. This is reasonable in the case of the author team because models like ANN and LSTM provide output probabilities, not just the final predicted class. Additionally, the ROC Curve is a graph that represents the classification capability of a model at various probability thresholds. In the case of Soft Voting, each sub-model makes predictions in the form of probabilities, and Soft Voting calculates the average of these probabilities to make the final prediction. When the ROC Curve is drawn based on these predictions, it reflects the classification capability not just at a fixed threshold but across a range of different thresholds - a softer and more flexible approach compared to Hard Voting, where there is only one fixed threshold (usually 50%).

Experimental Results of the Voting ANN Model

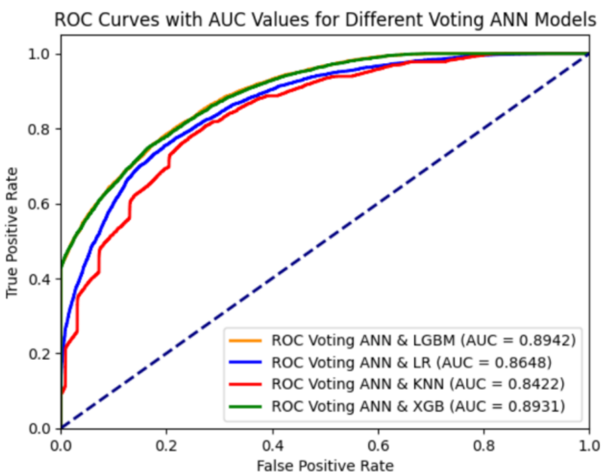
The results from the Voting ANN model indicate that the combination of algorithms like LR, KNN, LGBM, and XGBoost with ANN can create default prediction models with significant performance. Notably, LGBM + ANN and XGBoost + ANN stand out with high accuracy and ROC scores, demonstrating their good ability to distinguish between default and non-default cases. As illustrated in Table 6 and Figure 5, a high F1 score, along with balanced Precision and Recall, indicates that these models are not only accurate in their predictions but also minimize the number of false predictions.

Table 6
Experimental Results of Various Voting ANN Models

Voting ANN	LR + ANN	KNN + ANN	LGBM + ANN	XGBoost + ANN
Accuracy	80.90%	77.31%	84.99%	85.59%
Precision	73.47%	69.90%	78.77%	80.12%
Recall	77.49%	74.60%	77.23%	76.72%
F1-Score	74.95%	71.26%	77.95%	78.18%
ROC	86.48%	84.22%	89.42%	89.31%

Source: Author's work

Figure 5
The results of the ROC Curve for the Voting ANN model



Source: Author's work

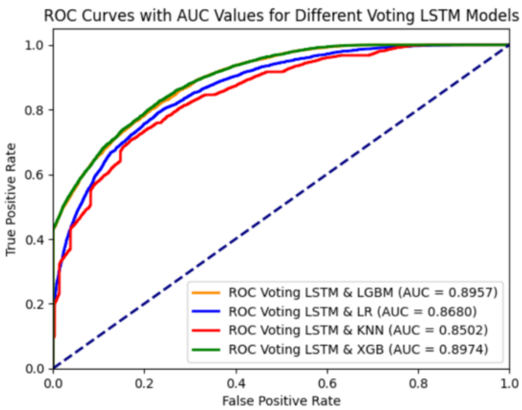
Experimental Results of the Voting LSTM Model
As illustrated in Table 7 and Figure 6, the Voting LR + LSTM and KNN + LSTM models show a slight improvement in accuracy and precision compared to the versions combined with ANN, possibly due to LSTM's ability to better capture sequential dependencies in complex data. Similarly, when comparing LGBM and XGBoost, the models combined with LSTM exhibit slightly higher accuracy and precision than their ANN counterparts. This suggests that LSTM may be a more suitable choice when paired with complex algorithms like LGBM and XGB.

Table 7
Experimental Results of Various Voting LSTM Models

Voting LSTM	LR + LSTM	KNN + LSTM	LGBM + LSTM	XGBoost + LSTM
Accuracy	81.60%	80.48%	85.60%	86.01%
Precision	74.08%	72.46%	80.08%	81.26%
Recall	77.19%	74.60%	76.89%	76.38%
F1-Score	75.34%	73.38%	78.27%	78.36%
ROC	86.80%	85.02%	89.57%	89.74%

Source: Author's work

Figure 6
The results of the ROC Curve for the Voting LSTM model



Source: Author's work

Summary of Experiments
In our experiments, voting and stacking ensembles that combined LGBM and XGBoost with LSTM or neural networks consistently outperformed one another, obtaining F1-Scores of approximately 78% in terms of loan default risk prediction. Stacking models demonstrated higher precision, indicating they were better at correctly identifying borrowers who default on their loans. On the other hand, voting models performed better in terms of recall and were better at identifying more defaulting debtors. F1-Scores for voting models were 0.1% higher than those for stacking models. Moreover, both models achieved high Area Under the ROC Curve values of nearly 90%, reflecting strong overall classification performance.

The decision to choose between these models depends on the specific objectives of the risk assessment. If minimizing false positives - avoiding misclassification of borrowers as defaulters when they will actually repay – is crucial, stacking models with higher precision is preferable. This reduces the risk of denying credit to reliable borrowers. Conversely, if the goal is to minimize false negatives by identifying as many defaulting borrowers as possible, voting models with higher recall are better suited, helping to mitigate potential losses from defaults.

Each approach has limitations. Due to their hierarchical structure, stacking models are more complex and computationally intensive, leading to longer training times. Voting models, while generally simpler and faster, might not capture complex interactions between base models as effectively. These findings highlight the importance of algorithm selection and ensemble techniques in improving loan default risk predictions. Balancing precision and recall, along with computational constraints, is essential for optimizing model performance in practical applications.

Conclusion

This study aims to enhance loan default prediction using an advanced ensemble learning technique that uses deep learning and machine learning models. Four algorithms, including LR, KNN, LGBM, and XGBoost, were combined with ANN and LSTM to improve the accuracy and reliability of default risk estimations for financial organizations.

To demonstrate the effectiveness of our models, we investigated and compared our results with several previous research models that used the Lending Club dataset. (Jin & Zhu, 2015) had their impressive single-model results with MPL and SVM up to 71.24% and 72.05% accuracy respectively. In addition, (Chen, Leu, Huang, Wang, & Takada, 2021) achieved impressive results, reaching approximately 77.9% accuracy by employing sampling methods such as SMOTE, Tomek links, and random sampling, in combination with regression and neural networks, particularly cost-sensitive learning models. One more research that included in our investigation came from (Xia, He, Li, Liu, & Ding, 2020). Their research demonstrated remarkable accuracy on the Lending Club dataset (2011-2013) by utilizing a clustering approach combined with decision tree-based methods, including CatBoost, RF, and LR, achieving accuracies ranging from 70% to nearly 80%. Our study consolidates previous results and presents improvements through the use of Stacking and Voting models, so combining the advantages of many core models. With this strategy, we were able to use the distinct benefits of every model, which resulted in notable gains in performance. Interestingly, a few combinations reached accuracies of about 85%; the XGB-LSTM model was particularly successful, with an astounding 86% accuracy. This illustrates how different approaches might be combined to maximize predicted results.

The results of our experiments clearly indicate the potency of ensemble approaches in improving predictive performance regarding risk of defaults in loan estimations. High F1-scores and high AUC values demonstrated the excellent discrimination ability between those potential payees and those at risk of defaulting for both. The stacking ensembles performed pretty well, reducing false positives - a critical component in precisely identifying trustworthy borrowers. Voting ensembles, on the other hand, proved to be better in recall, meaning that they are very good at recognizing a larger percentage of possible defaulters and reducing financial losses.

While our research focuses on predicting loan defaults using stacking and voting ensemble methods, we understand that our selection of algorithms is not exhaustive. Alternative models or combinations could be better suited to this complex task. In future studies, we will consider exploring other algorithms to address the intricacies of loan default prediction fully. In addition, explainable AI approaches, such as SHAP values or LIME, simplify model understanding beyond just targeting prediction metrics. This clarity is essential for compliance and building the confidence of professionals who utilize the models to make crucial decisions in the finance sector.

Acknowledgements: This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number B2023-34-04.

References

1. Abdelmoula, A. K. (2015). Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks. *Accounting and Management Information Systems, 14*(1), 79-106. https://doi.org/10.1109/OCIT59427.2023.10431007
2. Abhiram, P., Artham, N., Reddy, N., & Kumari, K. V. (2023). Predicting the borrower's genuineness in loan repayment through big data analytics. *2023 OITS International Conference on Information Technology (OCIT)* (pp. 767-774). IEEE: Piscataway, NJ, USA. https://doi.org/10.1109/OCIT59427.2023.10431007
3. Acito, F. (2023). k nearest neighbors. In F. Acito (Ed.), *Predictive analytics with KNIME: Analytics for citizen data scientists* (pp. 209-227). Cham, Switzerland: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45630-5_10
4. Adedapo, K. D. (2007). Analysis of default risk of agricultural loan by some selected commercial banks in Osogbo, Osun State, Nigeria. *International Journal of Applied Agriculture and Apiculture Research*, *4*(1&2), 24-29.
5. Alaloul, W. S., & Qureshi, A. H. (2020). Data processing using artificial neural networks. In D. Harkut (Ed.), *Dynamic data assimilation: Beating the uncertainties* (pp. 81–107). IntechOpen. https://doi.org/10.5772/intechopen.91935
6. Ali, A., Hamraz, M., Gul, N., Khan, D. M., Aldahmani, S., & Khan, Z. (2023). A k nearest neighbour ensemble via extended neighbourhood rule and feature subsets. *Pattern Recognition, 142*(1), 109641. https://doi.org/10.1016/j.patcog.2023.109641
7. Basha, S. A., Elgammal, M. M., & Abuzayed, B. M. (2021). Online peer-to-peer lending: A review of the literature. *Electronic Commerce Research and Applications, 48,* 101069. https://doi.org/10.1016/j.elerap.2021.101069
8. Brownlee, J. (2016). XGBoost with Python: *Gradient boosted trees with XGBoost and scikit-learn*. S.l.: Machine Learning Mastery. https://machinelearningmastery.com/xgboost-with-python/
9. Chen, D., Ye, J., & Ye, W. (2023). Interpretable selective learning in credit risk. *Research in International Business and Finance, 65*(C), 101940. https://doi.org/10.1016/j.ribaf.2023.101940
10. 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* (pp. 785-794). New York: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
11. Chen, Y. R., Leu, J. S., Huang, S. A., Wang, J. T., & Takada, J. I. (2021). Predicting default risk on peer-to-peer lending imbalanced datasets. *IEEE Access, 9*, 73103-73109. https://doi.org/10.1109/ACCESS.2021.3079701
12. Chi Tin. (2023, 07 26). *Ministry of Finance Makes a Breakthrough in Administrative Reform and Digital Transformation*. Retrieved from Ministry of Finance of Vietnam: https://mof.gov.vn/webcenter/portal/ttncdtbh/pages_r/l/chi-tiet-tin?dDocName=MOFUCM278175
13. Vietnam Governance. (2010, 6 16). Law No. 47/2010/QH12 by the *National Assembly*: LAW ON CREDIT INSTITUTIONS. Retrieved from Government Document System of Vietnam: https://vanban.chinhphu.vn/default.aspx?pageid=27160&docid=96074
14. Dhruv, C., Paul, D., Kumar, M. H., & Reddy, M. S. (2023). Framework for bank loan re-payment prediction and income prediction. *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 833-840). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/ICSCCC58608.2023.10176363

15. Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science, 14*, 241-258. https://link.springer.com/article/10.1007/s11704-019-8208-z
16. Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data, 11*(1), 23. https://doi.org/10.1186/s40537-024-00882-0
17. Fan, S. (2023). Design and implementation of a personal loan default prediction platform based on LightGBM model. *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)* (pp. 1232-1236). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/ICPECA56706.2023.10076254
18. Fang, J., & Ji, Z. (2024). Application of machine learning in loan default prediction. *Mathematical Modeling and Algorithm Application, 2*(2), 33-35. https://doi.org/10.54097/75k4fe13
19. Fauzi, M. A., & Yuniarti, A. (2018). Ensemble method for indonesian twitter hate speech detection. *Indonesian Journal of Electrical Engineering and Computer Science, 11*(1), 294-299. http://doi.org/10.11591/ijeecs.v11.i1.pp294-299
20. George, N. (2021, 2 1). All Lending Club loan data. Retrieved from *Kaggle*: https://www.kaggle.com/datasets/wordsforthewise/lending-club/data
21. Gupta, A., Pant, V., Kumar, S., & Bansal, P. K. (2020). Bank Loan Prediction System using Machine Learning. *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)* (pp. 423-426). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/SMART50582.2020.9336801
22. Hakkal, S., & Lahcen, A. A. (2024). XGBoost to enhance learner performance prediction. *Computers and Education: Artificial Intelligence, 7*, 100254. https://doi.org/10.1016/j.caeai.2024.100254
23. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation,* 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
24. Jayaram, E. S., Balachandar, G., & Kumar, K. (2024). Machine learning-based loan default prediction: Models, insights, and performance evaluation in peer-to-peer lending platforms. *Educational Administration: Theory and Practice, 30*(5), 12975-12989. http://dx.doi.org/10.53555/kuey.v30i5.5637
25. Jin, Y., & Zhu, Y. (2015). A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending. *2015 Fifth International Conference on Communication Systems and Network Technologies* (pp. 609-613). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/CSNT.2015.25
26. Kabari, L. G., & Onwuka, U. C. (2019). Comparison of bagging and voting ensemble machine learning algorithm as a classifier. *International Journals of Advanced Research in Computer Science and Software Engineering,* 9(3), 19-23.
27. Kalule, R., Abderrahmane, H. A., Alameri, W., & Sassi, M. (2023). Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs. *Scientific Reports, 13*(1), 9855. https://doi.org/10.1038/s41598-023-36096-2
28. Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., . . . Liu, T. (2017, 12). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Retrieved from *Microsoft Research1*: https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/
29. Kim, H., Cho, H., & Ryu, D. (2020). Corporate Default Predictions Using Machine Learning: Literature Review. *Sustainability, 12*(16), 6325. https://doi.org/10.3390/su12166325
30. Koç, U., & Sevgili, T. (2020). Consumer loans' first payment default detection: A predictive model. *Turkish Journal of Electrical Engineering and Computer Sciences, 28*(1), 167-181. https://doi.org/10.3906/elk-1809-190

31. Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering, 2*, 40-46. https://doi.org/10.1016/j.ijcce.2021.01.001
32. Li, F., Zhang, L., Chen, B., Gao, D., Cheng, Y., Zhang, X., . . . Huang, Z. (2020). An optimal stacking ensemble for remaining useful life estimation of systems under multi-operating conditions. *IEEE Access, 8*, 31854-31868. https://doi.org/10.1109/ACCESS.2020.2973500
33. Li, S., Ma, K., Niu, X., Wang, Y., Ji, K., Yu, Z., & Chen, Z. (2019). Stacking-based ensemble learning on low dimensional features for fake news detection. *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2730-2735. https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00383
34. Machado, M. R., Karray, S., & De Sousa, I. T. (2019). LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. pp. 1111-1116). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/ICCSE.2019.8845529
35. Pandey, D., & Pandey, B. K. (2022). An efficient deep neural network with adaptive galactic swarm optimization for complex image text extraction. In (Eds), V. Yadav, A. K. Dubey, H. P. Singh, G. Dubey, & E. Suryani, *Process mining techniques for pattern recognition* (pp. 121-137). Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781003169550-10
36. Qi, X. (2023). Factors influence loan default–A credit risk analysis. *International Conference on Economic Management and Green Development* (pp. 849-862). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0523-8_79
37. Rincy, T. N., & Gupta, R. (2020). Ensemble learning techniques and its efficiency in machine learning: A survey. *2020 2nd International Conference on Data, Engineering and Applications (IDEA)* (pp. 1-6). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/IDEA49133.2020.9170675
38. Sain, K., & Kumar, P. C. (2022). An Overview of Artificial Neural Networks. In K. Sain, & P. C. Kumar, *Meta-Attributes and Artificial Networking: A New Tool for Seismic Interpretation* (pp. 73-93). Hoboken, New Jersey: John Wiley & Sons. https://doi.org/10.1002/9781119481874
39. Satpute, S., Jayabalan, M., Kolivand, H., Assi, J., Aldhaibani, O. A., Liatsis, P., & Mahyoub, M. (2022). Loan default forecasting using StackNet. *The International Conference on Data Science and Emerging Technologies* (pp. 434-447). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-0741-0_31
40. Schonlau, M. (2023). Logistic regression. *In M. Schonlau, Applied statistical learning: With case studies in Stata* (pp. 49-71). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-031-33390-3_4
41. Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *Australasian joint conference on artificial intelligence* (pp. 1015-1021). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11941439_114
42. Thorat, M., Pandit, S., & Balote, S. (2022). Artificial neural network: A brief study. *Asian Journal for Convergence in Technology (AJCT), 8*(3), 12-16. https://doi.org/10.33130/AJCT.2022v08i03.003
43. Uddin, N., Ahamed, M. K., Uddin, M. A., Islam, M. M., Talukder, M. A., & Aryal, S. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application. *International Journal of Cognitive Computing in Engineering, 4*(6), 327-339. https://doi.org/10.1016/j.ijcce.2023.09.001

44. Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientific Reports, 12*(1), 6256. https://doi.org/10.1038/s41598-022-10358-x
45. Wang, C., Han, D., Liu, Q., & Luo, S. (2018). A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM. *IEEE Access, 7*, 2161-2168. https://doi.org/10.1109/ACCESS.2018.2887138
46. Wang, W., Zuo, X., & Han, D. (2024). Predict credit risk with XGBoost. *Applied and Computational Engineering, 74*(1), 164-177. https://doi.org/10.54254/2755-2721/74/20240462
47. Wolpert, D. H. (1992). Stacked generalization. *Neural networks, 5*(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1
48. Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting loan default in peer-to-peer lending using narrative data. *Journal of Forecasting, 39*(2), 39(2), 260-280. https://doi.org/10.1002/for.2625
49. Yadav, D., Sahoo, L., Mandal, S. K., Ravivarman, G., Vijayaraghavan, P., & Prasad, B. (2023). Using long short-term memory units for time series forecasting. *2023 2nd International Conference on Futuristic Technologies (INCOFT)* (pp. 1-6). Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/INCOFT60753.2023.10425756
50. Zhou, Y. (2023). Loan default prediction based on machine learning methods. *Proceedings of the 3rd International Conference on Big Data Economy and Information Management (BDEIM 2022)*. Zhengzhou, China: EAI. http://doi.org/10.4108/eai.2-12-2022.2328740

About the authors

Hoanh-Su Le (Ph.D.) received a B.E. in electronics and telecommunication, an MSc in MIS, and an MBA degree from Vietnam National University, HCM City, in 2009 and 2011. He received a Ph.D. degree in E-business from Pukyong National University, Republic of Korea, in 2016. Since 2011, he has been a faculty member. He is currently the dean (Tenure) of the Faculty of Information Systems at the University of Economics and Law, Vietnam National University, Ho Chi Minh City (VNU HCM). His research interests are data analytics, artificial intelligence, and fintech. Hoanh-Su Le is the author and can be contacted at: **sulh@uel.edu.vn**

Le Quang Chan Phong is a student in business analytics and artificial intelligence at the University of Economics and Law, Vietnam National University, Ho Chi Minh City. He is a research assistant at the Faculty of Information Systems, University of Economics and Law. His research interests are data processing, data analytics, and artificial intelligence. The author can be contacted at: **phonglqc20406c@st.uel.edu.vn**

Truong Cong Vinh is a student in business analytics and artificial intelligence at the University of Economics and Law, Vietnam National University, Ho Chi Minh City. He is a research assistant at the Faculty of Information Systems, University of Economics and Law. His research interests are data processing, data analytics, and artificial intelligence. The author can be contacted at: **vinhtc20416c@st.uel.edu.vn**

Ho Mai Minh Nhat is a student in business analytics and artificial intelligence at the University of Economics and Law, Vietnam National University, Ho Chi Minh City. He is a research assistant at the Faculty of Information Systems, University of Economics and Law. His research interests are data processing, data analytics, and artificial intelligence. The author can be contacted at: **nhathmm20416c@st.uel.edu.vn**

Jong-hwa Lee (Ph.D.) received a master's and doctorate degrees in business administration from Pukyong National University in 2014 and 2017, respectively. Since 2019, he has been a professor in the Department of E-Business at Dong-Eui University, Busan City, South Korea. His research interest areas are big data, artificial intelligence, digital finance, and data mining. Jong-Hwa Lee is the corresponding author and can be contacted at: **jhlee6050@deu.ac.kr**