



Research article

Optimizing corporate bankruptcy prediction through stacking ensemble of KNeighbors, decision tree, and random forest algorithms: Insights from Vietnamese enterprises

Nhat Minh Nguyen and Anh Hoang Le*

Faculty of Banking, Ho Chi Minh University of Banking, 36 Ton That Dam Street, District 1, Ho Chi Minh City, Vietnam

* **Correspondence:** Email: anhhlh_vnc@hub.edu.vn; Tel: +84-973-624-322.

Abstract: This study aimed to explore the prediction of corporate bankruptcy probability by applying the stacking ensemble approach that integrates KNeighbors, decision tree, and random forest algorithms, comparing the result with the application of each of the methods individually. The dataset of 6819 firms, monitored by Vietnamese commercial banks from 2014 to 2022, is based on 94 financial features extracted from the financial statements of the firms. The study provides empirical support for the stacking ensemble approach consisting of three interpretable machine learning algorithms: KNeighbors, decision tree, and random forest, outperforming the application of each of the methods individually for the prediction of the default probability of the firm. The stacked approach yielded significant improvements in the accuracy, precision, specificity, and F1 score of the prediction. The validation of the methods was based on a confusion matrix and the Wilcoxon signed-rank test. This study contributes to the literature by providing applied empirical validation within two relatively underexplored areas. First, while the majority of previous research on bankruptcy prediction has focused on the effectiveness of more complex algorithms, such as boosted ensemble methods and neural networks, the effectiveness of less complex models within ensemble methods has received little attention. The study provides evidence of the effectiveness of K-nearest neighbors, decision tree, and random forest models within a stacking ensemble method. Second, while previous research has relied on data from developed markets, the effectiveness of financial risk assessment models within emerging markets has received little attention. The paper provides evidence of the effectiveness of a stacking ensemble method within an emerging market, namely Vietnam, where financial characteristics of companies are significantly different from those found in developed markets.

Keywords: corporate bankruptcy prediction; ensemble stacking models; financial risk management; machine learning; feature importance analysis

JEL Codes: E44, F31, F37, G15

1. Introduction

In today's economy, corporate bankruptcy is an extremely important issue, both theoretically and practically. If a company cannot meet its financial obligations on time, it may be forced to leave the market through a declaration of bankruptcy. The bankruptcy process causes financial difficulties for the company and poses risks to both investors and the economy. Therefore, it is crucial for stakeholders to understand the causes of corporate bankruptcy in order to apply effective risk management measures for their companies. The most important causes of corporate bankruptcy risk are often related to macroeconomic factors such as interest rates, inflation, and economic growth. These factors can exacerbate difficulties for companies on the verge of bankruptcy, especially in unfavorable economic environments (Issa et al., 2024; Ghazalat and AlHallaq, 2024). High borrowing costs can also exacerbate financial difficulties for a company (Sandin and Porporato, 2007).

In recent years, numerous theoretical and practical studies have utilized machine learning models to predict business failures. Specifically, K-nearest neighbors (KNN), decision trees, random forests, and stacking models have been used to estimate the probability of business failure. The results of these studies show that KNN models are effective in identifying nonlinear patterns. However, fine-tuning the model to minimize overfitting when using KNNs is a difficult problem, especially when working with multidimensional datasets (Chahboun & Maaroufi, 2022). Decision trees have also been used to estimate the probability of business failure. This model is particularly attractive for predicting business failure because it can be represented hierarchically, making the results easier to understand. However, it is prone to overfitting, a problem that has been addressed through the combination with other models such as random forest or stacking (Uddin et al., 2023). Unlike KNN, the random forest model has been shown to be effective when working with multidimensional datasets characterized by nonlinear financial indicators (Kuyoro et al., 2022). Evidence suggests that it performs better than resampling-based models like SMOTE when working with unbalanced datasets related to corporate bankruptcy (Muslim et al., 2024). The stacking model has also been used to estimate the probability of corporate bankruptcy, with the ability to identify nonlinear patterns in financial indicators related to corporate bankruptcy (Amirshahi & Lahmiri, 2024; Choi et al., 2018). This model has also proven effective when working with different data formats, allowing for workflow implementation and making the model easier to understand (Chen et al., 2020; Jabeur et al., 2021; Climent et al., 2019).

Machine learning models have also been increasingly used for bankruptcy prediction. However, most of these studies focus on more complex machine learning models like XGBoost, LightGBM, and neural networks (Nguyen and Ngo, 2025). There is a tendency to ignore simpler machine learning models like K-nearest neighbors and decision trees with ensemble models (Mienye and Sun, 2022). Besides this, even though stacking models have yielded promising results, only a small number of studies have evaluated their performance with only interpretable models and for emerging markets. A major research gap is identified with regard to whether these commonly used interpretable models, integrated with a stacking model, perform better than even a single predictive model. Furthermore,

most of these studies have been conducted in developed economies, which is a limitation for emerging markets like Vietnam because of differences in corporate characteristics.

Financial ratios have also been commonly used for bankruptcy prediction. Researchers have also studied various groups of ratios like profitability ratios, liquidity ratios, leverage ratios, efficiency ratios, growth ratios, per-share ratios, cash flow ratios, solvency ratios, and expense ratios to identify early warning signs for corporate bankruptcy. Higher return on assets and net income to equity reduce bankruptcy probability (He et al., 2022). Strong performance with regard to liquidity ratios (Bose et al., 2021) and strong asset turnover ratios (Demerjian et al., 2012) also signal strong corporate health. Higher leverage ratios increase insolvency risk for companies (Graham et al., 2015), and poor cash flow ratios often signal corporate crisis (Richardson et al., 2005). Despite their widespread usage for bankruptcy prediction, researchers continue to argue about their relative importance in different economic contexts (Beaver et al., 2011).

Therefore, this study proposes a stacking ensemble model that integrates three algorithms—KNeighbors, decision tree, and random forest—with the aim of improving the performance of corporate bankruptcy prediction. In addition, the study identifies key financial indicators that contribute to interpreting the model's predictive outcomes. Using a dataset comprising 6819 Vietnamese corporations collected between 2014 and 2022, the study evaluates model performance using six confusion matrix metrics: accuracy, specificity, sensitivity, precision, F1 score, and area under the curve (AUC). Furthermore, to ensure the reliability of the model's superiority over individual models, the Wilcoxon signed-rank test is also employed.

The current study makes the following contributions to applied empirical research in the relevant literature: First, although the stacking ensemble method is a well-established technique in machine learning research, its application with interpretable machine learning methods such as KNeighbors, decision tree, and random forest, has not been empirically well investigated in bankruptcy prediction problems. Second, the current research extends the relevant empirical research in an emerging economy such as Vietnam, in which the financial characteristics of corporations are different from those in developed economies. Finally, this research contributes to the relevant literature by finding the most significant financial ratios that affect the results of bankruptcy prediction problems using the stacking ensemble method. Overall, the current research makes significant empirical contributions in the relevant field, aiming to extend or validate existing research methods rather than develop novel ones.

The remainder of the study includes a literature review, a detailed research methodology comprising the research design and algorithm implementation, followed by the presentation of results and discussion, and finally, a conclusion that highlights practical implications and directions for future research.

2. Literature review

A systematic literature search was conducted to identify relevant studies on corporate bankruptcy prediction using machine learning methodologies. Multiple academic databases were searched, including Web of Science, Scopus, IEEE Xplore, ScienceDirect, and Google Scholar. The search employed the following query combinations: (“corporate bankruptcy” OR “financial distress” OR “bankruptcy prediction”) AND (“machine learning” OR “ensemble learning” OR “stacking ensemble”); (“KNeighbors” OR “K-Nearest Neighbors” OR “Decision Tree” OR “Random Forest”) AND (“bankruptcy prediction” OR “credit risk”); (“stacking” OR “ensemble methods”) AND (“emerging markets” OR “Vietnam” OR “developing economies”); (“financial ratios” OR “financial

indicators”) AND “bankruptcy prediction”. Studies were included if they (1) were published between 2014 and 2024, (2) addressed corporate bankruptcy or financial distress prediction, (3) employed machine learning or ensemble learning techniques, (4) appeared in peer-reviewed journals or conference proceedings, and (5) were available in English.

The area of research on the prediction of corporate bankruptcy is of great importance, as it has significant impacts on financial risk management. While traditional models like logistic regression are effective to some extent, more complex methods are needed for particularly intricate financial data. The aforementioned recent innovations in machine learning models, including the implementation of KNeighbors, decision trees, random forests, and stacking, have greatly improved the accuracy and reliability of bankruptcy forecasts. Nevertheless, stacking methods have demonstrated the ability to substantially enhance prediction accuracy and adapt through the synergy of various base models, overcoming the limitations of others (Wang and Chi, 2024; Iparraguirre-Villanueva and Cabanillas-Carbonell, 2024).

The debate on the effectiveness of various machine learning models in predicting bankruptcy reveals diversity, with some studies indicating that different models perform best in individual cases. Random forest is often cited as a prominent option due to its power to handle large datasets and mitigate overfitting. Breiman (2001) and Probst et al. (2019) acknowledged that random forest strikes a balance between complexity and efficiency, while Biau et al. (2019) highlighted its robustness and versatility.

Despite the recognition that random forest receives, some researchers favor stacking models for their ability to integrate multiple algorithms. Wolpert, who introduced stacked generalization in 1992 (Wolpert, 1992), together with Chen et al. (2020), advocated for stacking in contexts where no single model captures all data complexities. Nagavelli et al. (2022) confirmed its advantages, particularly in the banking sector.

As discussed in the Introduction, it is a well-documented phenomenon in the machine learning literature that the overall performance of classification models can be significantly enhanced by the application of ensemble methods. However, the overall performance of some of the most basic yet effective classification algorithms in a stacking framework for bankruptcy prediction has not been extensively analyzed in the existing literature. While some previous studies have focused on more complex yet effective classification algorithms like gradient boosting machines and deep neural networks (Carmona et al., 2019), less attention has been devoted to the overall performance of stacking algorithms composed of only interpretable classifiers.

Prior research has also delineated the constraints of each methodology. KNeighbors might have trouble scaling up when working with data that has a lot of dimensions (Uddin et al., 2022; Zhang, Li, and Li, 2023). Decision trees, despite their interpretability, generally necessitate integration with ensemble methods to attain optimal performance (Broelemann and Kasneci, 2019; Carrizosa et al., 2021). In the meantime, stacking needs a lot of computer power, but random forest is still a good choice when computer power is limited (Barboza et al., 2017).

However, based on the available empirical results, the stacking approach improves the prediction of corporate bankruptcy by capitalizing on the combined characteristics of the base models. Analysis of the relative performance of the stacking models reveals that the KNeighbors algorithm, when used within the stacking architecture, can achieve better classification accuracy compared to the implementation of the decision trees and random forest models alone (Zhao et al., 2015; Wang & Chi, 2024). Wang et al. (2024) demonstrated the benefits of using stacked generalization models based on the Bayesian ridge regression algorithm, which have improved the accuracy of the models compared

to a single algorithm. Noh (2023) showed the effectiveness of the stacking models in overcoming the problem of class imbalance, which is commonly observed in datasets used to predict corporate bankruptcy, while Jiang et al. (2020) quantified the benefits of using stacking models to capture the complex interactions of the financial variables. The combined implementation of the base models using the stacking architecture can capture both linear and nonlinear relationships.

Table 1. Summary of key studies on machine learning for corporate bankruptcy prediction.

Authors (year)	Model/method	Key findings
Ohlson (1980)	Solvency ratios	Critical for assessing long-term debt repayment ability and credit risk
Wolpert (1992)	Stacked generalization	Introduced an ensemble method integrating multiple algorithms
Breiman (2001)	Random forest	Balances complexity and efficiency in bankruptcy prediction
Anderson & Lanen (2002)	Expense ratios	Reflects cost-effectiveness and profitability through automated spending control
Richardson et al. (2005)	Cash flow ratios	Predicts future financial crises based on changes in financial situations
Beyer et al. (2010)	Per-share metrics (EPS)	Significant impact on stock prices and investor decisions
La Rocca et al. (2011)	Growth metrics	Indicates future growth potential and investor attractiveness
Beaver et al. (2011)	Overall financial indicators	Essential for assessing corporate health in predictive analytics
Graham et al. (2015)	Leverage ratios	High leverage increases bankruptcy risk through capital structure
Zhao et al. (2015)	KNeighbors in ensemble	Superior accuracy when integrated in an ensemble architecture
Barboza et al. (2017)	Random forest vs. stacking	Random forest preferable with limited computational resources
Biau, Cadre, & Rouvière (2019)	Random forest	Demonstrates robustness and versatility
Broelemann & Kasneci (2019)	Decision tree	Requires ensemble integration for optimal performance
Probst et al. (2019)	Random forest	Confirms balance between complexity and efficiency
Chen et al. (2020)	Stacking	Effective when no single model captures all data complexities
Jiang et al. (2020)	Ensemble methods	Captures complex interactions among financial variables
Carrizosa et al. (2021)	Decision tree	Necessitates ensemble methods to improve accuracy
Nagavelli, et al. (2022)	Stacking	Confirms advantages in banking sector applications
Uddin et al. (2022)	KNeighbors	Scalability issues with high-dimensional data
Noh (2023)	Stacking	Effectively mitigates class imbalance in bankruptcy datasets
Zhang, Li, & Li (2023)	KNeighbors	Challenges in high-dimensional feature spaces
Iparraguirre-Villanueva & Cabanillas-Carbonell (2024)	Stacking	Enhances prediction accuracy through model synergy
Wang & Chi (2024)	Stacking ensemble	KNeighbors achieves superior performance in an ensemble framework
Wang et al. (2024)	Stacking with Bayesian ridge regression	Significant improvements using the meta-learner approach

Along with model-based methods, financial ratios have been used for a long time to show early indicators of bankruptcy. Researchers have examined many sets of ratios to detect financial trouble. For instance, leverage ratios show how a company's capital is structured. Too much leverage can mean a higher chance of bankruptcy, as Graham et al. (2015) pointed out. Growth-related metrics indicate

the company's potential for future growth and its attractiveness to investors (La Rocca et al., 2011). Earnings per share (EPS), a crucial per-share indicator, is recognized to have a big effect on stock prices and investor decisions (Beyer et al., 2010). Cash flow ratios, which illustrate how well a company can handle changes in its financial situation, have been found to forecast future financial crises (Richardson et al., 2005). Solvency ratios are significant for figuring out credit risk and the chance of bankruptcy, since they show how well a company can survive and pay off its debts over the long run (Ohlson, 1980). Expense ratios, on the other hand, show how cost-effective a business is, and automated spending control is a big part of making money (Anderson and Lanen, 2002). Despite disagreements about how important these ratios are, most scholars agree that financial indicators are very important for figuring out how healthy a company is and for predictive analytics (Beaver et al., 2011).

Synthesizing the aforementioned literature highlights two key observations. First, a number of studies have demonstrated the promise of ensemble-based approaches, with a particular emphasis on stacking architectures, to corporate bankruptcy prediction, although the assessment of such models with only interpretable base classifiers is surprisingly understudied. Second, the majority of the empirical evidence supporting such models has been derived from developed economies, which has a number of implications for the generalizability of such research to emerging markets with divergent corporate financial profiles. Accordingly, this study's research design was developed to address this gap through the creation and evaluation of a stacking ensemble model with three interpretable classifiers, which was deployed over a dataset of Vietnamese firms between 2014 and 2022. Table 1 summarises the key studies reviewed in this section, together with the models they employ and their principal findings.

3. Methodology

3.1. Research design

This research implements a stacking ensemble architecture for corporate bankruptcy prediction, integrating three interpretable machine learning algorithms: K-nearest neighbors, decision tree, and random forest. The methodology comprises five sequential phases as depicted in Figure 1: data preprocessing, feature transformation, model construction, performance assessment, and statistical validation.

Data preprocessing ensures dataset integrity and quality through systematic removal of duplicate records and imputation of missing values to maintain consistency across financial indicators. Class imbalance, a prevalent challenge in bankruptcy prediction datasets, is addressed using BorderlineSMOTE—a synthetic oversampling technique that enhances minority class (bankrupt firms) representation by generating synthetic samples near decision boundaries while proportionally reducing majority class dominance.

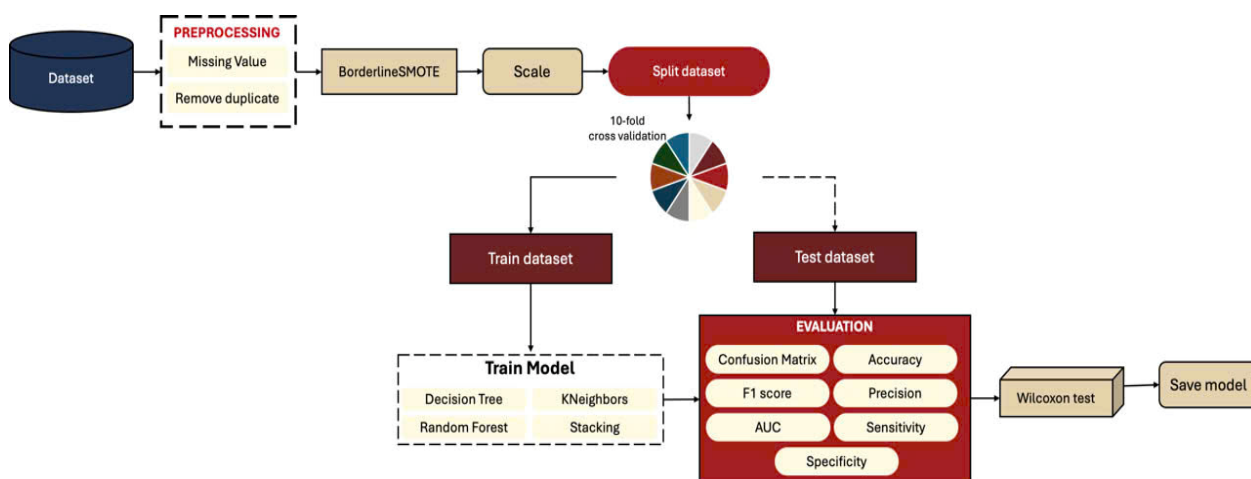


Figure 1. Ensemble model training process for predicting corporate bankruptcy.

Next, a scaling method is used to normalize the data such that all of the variables have the same effect on the model training. Then, 10-fold cross-validation splits the dataset into training and testing sets. This makes the model's performance more reliable and less biased when it is being tested. In order to reduce information leakage during the stacking process, a nested cross-validation approach was used. In each of the ten folds of the outer cross-validation process, an inner five-fold cross-validation approach was used solely for the purpose of generating the meta-features. The base classifiers, namely K-nearest neighbors, decision tree, and random forest, were trained on four of the five folds and were used to make predictions on the remaining held-out fold. These predictions were then concatenated to yield the complete set of meta-features, which were then used to train the meta-learner (logistic regression). The key factor to note is that the test set was completely isolated from all model training procedures, i.e., none of the base classifiers and the meta-learner were ever provided with any information regarding the test set during their respective training procedures. This approach to data partitioning is a critical factor in establishing the validity of the results obtained with the stacking ensemble model, which will be discussed further in the following sections.

With regard to feature selection and dimensionality reduction, this study has maintained all 94 financial variables throughout the modeling process for a number of methodological and operational reasons. First, tree-based classifiers, such as decision tree and random forest, perform feature selection implicitly through their recursive splitting mechanisms, which assign lower weights to certain features that possess lower discriminatory power without requiring an explicit feature selection mechanism (Breiman, 2001). Second, although the K-nearest neighbors algorithm is adversely affected by feature spaces with a large number of dimensions, this issue is circumvented to a large extent owing to the normalization process used during data preprocessing, which normalizes all variables to a comparable numerical scale.

Third, and most importantly from a practical point of view, the retention of all financial ratios mirrors actual credit risk assessment processes in Vietnamese commercial banks, in which risk analysts typically examine the entire set of financial ratios rather than a reduced set of variables. This choice of design thus contributes to the practical applicability of the research results obtained in this investigation.

Four algorithms are trained during the model building phase: KNeighbors, decision tree, random forest, and a stacking ensemble that merges the three foundation models. The training set is used to

teach each model, and the test folds are used to see how well it can predict. The stacking model uses a meta-learner to combine the outputs of the base learners, which makes the overall predictions more accurate.

This study used common classification measures from the confusion matrix, like accuracy, precision, sensitivity, specificity, F1 score, and AUC, to evaluate the model. These metrics give a full picture of how accurate and strong the model is in both classes.

To assess the statistical significance of performance improvements, the Wilcoxon signed-rank test is applied to compare the stacking ensemble with its base models across all validation folds. Finally, the best-performing model is serialized and saved for future deployment. All computational experiments were implemented in the Python programming language (version 3.10) utilizing the following open-source software libraries: scikit-learn (version 1.3) for model construction, cross-validation, hyperparameter configuration, and performance metric computation; imbalanced-learn (version 0.11) for the implementation of the BorderlineSMOTE oversampling procedure; pandas (version 2.0) and NumPy (version 1.24) for data manipulation, cleaning, and numerical computation; and Matplotlib (version 3.7) in conjunction with Seaborn (version 0.12) for statistical visualization. The stacking ensemble was operationalized using scikit-learn's `StackingClassifier` with logistic regression designated as the final meta-learner, and 5-fold inner cross-validation was employed for the generation of meta-features from base learner predictions. Model serialization for potential operational deployment was conducted using the `joblib` library. All experiments were executed on a computing environment equipped with an Intel Core i7 processor (2.6 GHz) and 16 GB of RAM, operating under the Ubuntu 22.04 Linux distribution. The complete analytical pipeline—encompassing data preprocessing, model training, evaluation, and visualization—was developed and executed within Jupyter Notebook to facilitate procedural transparency and reproducibility.

3.2. The base classifiers and stacking ensemble model

KNeighbors (KNN)

KNN is a simple but powerful algorithm that uses a nonparametric method to find patterns and changes in data that are close to each other. It puts a new observation into the group of its closest points in the feature space that has the most points. This makes it easier to understand and use. Even though KNN is easy to understand, it can come close to finding complex decision limits. This makes it a good choice for grouping datasets that do not follow a normal distribution. KNN is a very flexible machine learning method because it can work with a wide range of data types by using distance functions. Choosing the k parameter (the number of neighbors to consider) carefully makes the process work best. Cross-validation is often used to find the best value for k so that the bias-variance tradeoff is balanced.

KNN is an example of an instance-based learning approach. In addition to the basic Euclidean distance, more advanced distance metrics like the Minkowski distance can be used, which lets the model adapt to additional norms:

$$d(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{\frac{1}{p}} \quad (1)$$

where p is a number that determines the distance measure. For instance, if $p = 1$, the distance is the Manhattan distance, and if $p = 2$, the distance is the Euclidean distance. The KNN is more beneficial because you can choose the distance metric, which means it may be used with different sorts of data distributions.

Weighted voting can be added to KNN to make it even better. In this system, neighbors who are closest to each other have more say in the categorization choice.

$$\hat{y} = \underset{y_j \in \{y_1, y_2, \dots, y_k\}}{\operatorname{argmax}} \sum_{i=1}^k \omega_i I(y_i = y_j) \quad (2)$$

where the weight w_i is inversely proportional to the distance $d(x, x_i)$, often calculated as

$$w_i = \frac{1}{d(x, x_i) + \epsilon} \quad (3)$$

with ϵ being a small constant to avoid division by zero. This weighted approach enhances KNN's ability to handle noisy data and make more refined predictions.

Decision tree (DT)

A DT is a hierarchical model built using binary recursive partitioning of data through a tree-based approach. Initially, the algorithm identifies the feature that best separates the classes by dividing the feature space into increasingly homogeneous regions. Decision trees are widely recognized for their interpretability, as the decision-making process is clearly displayed through the tree structure, making it easy to understand how predictions are made. Moreover, decision trees can capture both linear and nonlinear relationships within the data, which makes them highly adaptable. However, the model is prone to overfitting; therefore, pruning and limiting the tree depth are commonly used techniques to improve generalization.

A decision tree uses recursive binary partitioning, whose purpose is impurity reduction. The splitting criterion may be refined by using the gain ratio, which modifies information gain based on the inherent information of a split:

$$\text{Gain Ratio} = \frac{\text{Information Gain}}{\text{Split Information}} \quad (4)$$

where split information is defined as:

$$\text{Split Information} = - \sum_{i=1}^k \frac{N_i}{N} \log_2 \left(\frac{N_i}{N} \right) \quad (5)$$

Here, N_i is the number of samples in the i -th subset, and N is the total number of samples. This penalizes splits with many branches, reducing the tendency to favor splits that produce many small partitions.

For continuous features, the optimal split is found by minimizing the weighted impurity across potential split points t :

$$\min_t \left[\frac{n_{\text{left}}}{n} \text{Gini}(\text{left}(t)) + \frac{n_{\text{right}}}{n} \text{Gini}(\text{right}(t)) \right] \quad (6)$$

This advanced optimization ensures that decision trees are more precise in handling continuous variables, making them more effective in capturing complex relationships.

Random forest (RF)

Random forests are a type of combined machine learning method that makes predictions by combining predictions from multiple different decision trees. Each decision tree is constructed from

the dataset after selecting bootstrap patterns. Therefore, random forests help reduce variance and prevent overfitting, which frequently occurs with decision tree methods. This model performs strongly on multidimensional datasets and maintains accuracy even when variables in the model are noisy or irrelevant. It has strong generalizability on unobserved data and can be used for both classification and regression problems. In the financial field, random forests can be applied to assess financial risk.

The random forest makes a group of decision trees by using bootstrapping and choosing features at random. The final prediction is made by a majority vote:

$$\hat{y} = \underset{y_j \in \{\hat{y}_1(x), \hat{y}_2(x), \dots, \hat{y}_M(x)\}}{\operatorname{argmax}} \sum_{m=1}^M I(\hat{y}_m(x) = y_j) \quad (7)$$

For regression tasks, predictions are averaged:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M \hat{y}_m(x) \quad (8)$$

An advanced perspective on random forest includes the concept of out-of-bag (OOB) error:

$$OOB \text{ Error} = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{y}_{OOB}(x_i)) \quad (9)$$

where $\hat{y}_{OOB}(x_i)$ is the prediction made by aggregating the trees that do not include the i -th data point in their bootstrap sample, and $L(y_i, \hat{y})$ is the loss function (e.g., 0-1 loss for classification).

Additionally, random forest benefits from the bias-variance decomposition, which shows that the ensemble reduces variance without significantly increasing bias. Mathematically, the variance of the random forest is reduced as:

$$\operatorname{Var}(\hat{y}) \approx \frac{1}{M} \operatorname{Var}(y) + \left(1 - \frac{1}{M}\right) \operatorname{Cov}(\hat{y}_1, \hat{y}_2) \quad (10)$$

where M is the number of trees, and $\operatorname{Cov}(\hat{y}_1, \hat{y}_2)$ represents the covariance between individual tree predictions. The random feature selection process reduces the covariance between trees, further lowering the overall variance.

For classification tasks, the Gini importance (mean decrease in impurity) provides a measure of feature importance:

$$\text{Gini Importance}(f_i) = \sum_{t \in T} \Delta \text{Gini}(t, f_i) \quad (11)$$

where $\Delta \text{Gini}(t, f_j)$ represents the change in Gini impurity for feature f_j at node t , summed over all trees T in the forest. This allows the model to identify the most significant features, enhancing interpretability and performance.

Stacking ensemble model

Stacking is a combination method used to combine predictions made by multiple models to improve the accuracy of the individual models. In this study, we used stacking with three explainable machine learning models, including K-nearest neighbor, decision tree, and random forest. Predictions made by the component models were used to train a new model, called a meta-learner. The model used for stacking is typically a simple model, such as logistic regression, used to select the best combination

of predictions made by the component models (Nwogugu, 2006). This method can be expressed mathematically as follows:

$$\hat{y}(x) = \sum_{b=1}^B \beta_b \hat{y}_b(x) + \epsilon$$

where $\hat{y}_b(x)$ is the prediction from the base model b – *th*, β_b is the weight assigned to that model, and ϵ is the error term.

The combination in this stacking model allows leveraging the distinct strengths of each underlying model. Specifically, KNeighbors excel in identifying local patterns, decision trees provide high explanatory power, and random forests offer robust noise immunity and overfitting resistance. This creates a balance between predictive accuracy and model explanatory power, aligning with key principles in credit risk assessment practices at banks and financial institutions.

For all base learner models, the implementation was done using the default parameter settings as specified by the scikit-learn library (version 1.3), which are well-established parameter configurations in the machine learning literature (Pedregosa et al., 2011). These configurations are specified below. For the KNeighbors classifier, the value of k was set to 5, using the Minkowski distance metric with $p = 2$ (or the Euclidean distance) and uniform weightings of the neighbors, which provides a balance between computational efficiency and accuracy for the classifier on moderate-sized datasets. For the decision tree classifier, the Gini impurity criterion was used to select the attribute to be used to split the instances, with no maximum depth specified for the tree, thus allowing the tree to grow to the maximum depth until the leaves are pure or have fewer than two training instances in each node.

For the random forest classifier, an ensemble of 100 decision trees was constructed, each trained on a bootstrap sample of the training instances and using the Gini criterion to select the attribute to be used to split the instances, with the number of features to consider at each node set to the square root of the total number of features, which is a well-established parameter setting to promote diversity of the decision trees and prevent correlation between them (Breiman, 2001; Probst et al., 2019). For the stacking classifier, the meta-learner was implemented using a logistic regression classifier with L2 (ridge) regularization, using the default value of the regularization strength parameter C set to 1.0.

The deliberate adoption of default hyperparameter settings serves a dual methodological purpose. First, it establishes a standardized baseline comparison that isolates the architectural contribution of the stacking ensemble from the potentially confounding influence of algorithm-specific optimization, thereby enabling a cleaner assessment of the ensemble's inherent performance advantage over individual classifiers. Second, default configurations represent a pragmatic and immediately replicable starting point for practitioners at financial institutions who may lack the computational infrastructure required for exhaustive hyperparameter search procedures.

3.3. Classification performance metrics

Bankruptcy prediction models can be evaluated using a range of metrics derived from the confusion matrix, including accuracy, sensitivity, specificity, precision, F1 score, and the area under the curve (AUC). These metrics provide a comprehensive assessment of model performance, and through refinement and optimization procedures, they contribute to improving the model's ability to detect financial distress effectively

Confusion matrix

The confusion matrix is a fundamental evaluation tool for classification models, particularly in financial applications such as bankruptcy prediction. It provides detailed insights into both correct and incorrect predictions, helping to assess the model's overall reliability. Its importance lies in supporting informed decision-making, which is especially critical in the financial sector. As emphasized by Krauss and Stübinger (2017), a proper understanding and use of the confusion matrix can lead to better financial decisions and outcomes. Table 2 presents the confusion matrix, while Table 3 reports the associated performance metrics derived from it.

Table 2. Confusion matrix.

Predicted class			
Actual class	Classes	Non-bankruptcy	Bankruptcy
	Non-bankruptcy	True negative (TN): Cases where the model correctly predicts that customers will not go bankrupt, and they do not.	False positive (FP): Cases where the model incorrectly predicts bankruptcy, yet the customers do not actually go bankrupt.
	Bankruptcy	False negative (FN): Cases where the model fails to predict bankruptcy when the customers actually go bankrupt.	True positive (TP): Cases where the model accurately predicts that customers will go bankrupt, and they indeed do.

Table 3. Accuracy, sensitivity, specificity, precision, F1 score, AUC.

Metrics	Definition	Formula
Accuracy	Represents the proportion of total correct predictions (i.e., both true positives and true negatives) relative to the overall number of cases	Accuracy = $\frac{(TN + TP)}{(TN + FN + TP + FP)}$
Sensitivity	Assesses the model's capacity to accurately detect actual bankruptcies among corporations.	Sensitivity = $\frac{TP}{(TP + FN)}$
Specificity	Evaluates how effectively the model identifies corporations that do not go bankrupt	Specificity = $\frac{TN}{(TN + FP)}$
Precision	Measures the accuracy of bankruptcy predictions, calculated as the proportion of true positives among all corporations predicted to go bankrupt	Precision = $\frac{TP}{(TP + FP)}$
F1 score	Provides a harmonized measure of precision and sensitivity by taking into account both false positives and false negatives in its calculation	$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$
AUC	Assesses the entire area under the ROC curve, effectively summarizing the model's ability to discriminate between corporations that go bankrupt and those that do not, across various threshold settings	

To determine whether the stacking ensemble model significantly outperforms traditional machine learning classifiers in predicting corporate bankruptcy, the study formulates the following hypotheses for statistical testing:

Null hypothesis (H_0): There is no statistically significant difference in model performance between the stacking ensemble model and any of the base classifiers (KNeighbors, decision tree, and random forest) across the evaluated metrics such as accuracy, sensitivity, specificity, precision, F1 score, and AUC.

Alternative hypothesis (H_1): The stacking ensemble model demonstrates significantly better performance than at least one of the base classifiers across the aforementioned metrics.

Given that the data involve paired comparisons and may not follow a normal distribution, the study employs the Wilcoxon signed-rank test to perform pairwise comparisons between the stacking ensemble model and each baseline model. The level of statistical significance is set at $\alpha = 0.05$.

4. Research results

4.1. Data research

Table 4. Dataset group of variables.

	Group of variables	Description
Explanatory group of variables	Profitability ratios	Return on Assets Before Interest and Depreciation, Return on Assets After Tax, Return on Assets After Tax and Depreciation, Operating Gross Margin, Realized Sales Gross Margin, Operating Profit Rate, Pre-Tax Net Interest Rate, After-Tax Net Interest Rate, Non-Operating Income to Expenditure Ratio, Continuous After-Tax Interest Rate, Operating Profit per Share, Net Profit Before Tax per Share, Realized Sales Gross Profit Growth Rate, Operating Profit Growth Rate, After-Tax Net Profit Growth Rate, Regular Net Profit Growth Rate, Continuous Net Profit Growth Rate, Operating Profit to Capital, Net Profit Before Tax to Capital, Net Income to Total Assets, Gross Profit to Sales, and Net Income to Stockholders' Equity.
	Liquidity ratios	Current Ratio, Quick Ratio, Quick Assets to Current Liabilities, Cash to Current Liabilities, Current Liabilities to Assets, Quick Assets to Total Assets, Cash to Total Assets, Cash Flow to Sales, Current Asset Turnover Rate, and Quick Asset Turnover Rate.
	Leverage ratios	Interest Expense Ratio, Debt to Equity Ratio, Debt to Asset Ratio, Equity to Asset Ratio, Long-Term Fund Suitability Ratio, Borrowing Dependency, Contingent Liabilities to Equity Ratio, Liability to Equity, Degree of Financial Leverage, Interest Coverage Ratio, Equity to Liability, and Equity to Long-Term Liability.
	Efficiency ratios	Total Asset Turnover, Accounts Receivable Turnover, Average Collection Days, Inventory Turnover Rate, Fixed Asset Turnover Frequency, Equity Turnover Rate, Revenue per Employee, Operating Profit per Employee, Working Capital Turnover Rate, Cash Turnover Rate, and Total Assets to GNP Price.
	Growth ratios	Total Asset Growth Rate, Net Value Growth Rate, Total Asset Return Growth Rate, and Cash Reinvestment Ratio.
	Per share metrics	Book Value Per Share, Adjusted Book Value Per Share, Final Book Value Per Share, Persistent Earnings Per Share, Cash Flow Per Share, Revenue Per Share, and Net Profit Before Tax Per Share.
	Cash flow ratios	Cash Flow Rate, Cash Flow to Total Assets, Cash Flow to Liabilities, Cash Flow from Operations to Assets, and Cash Flow to Equity.
	Solvency ratios	Working Capital to Total Assets, Current Assets to Total Assets, Working Capital to Equity, Current Liabilities to Total Liabilities, Current Liability to Equity, Long-Term Liability to Current Assets, and Retained Earnings to Total Assets.
	Expense ratios	Operating Expense Rate, Research and Development Expense Rate, Total Income to Expense Ratio, and Total Expense to Assets Ratio.
	Miscellaneous	Interest-Bearing Debt Interest Rate, Effective Tax Rate, Allocation Rate per Employee, Operating Funds to Liability, Inventory and Receivables to Equity, Inventory to Working Capital, Inventory to Current Liability, Liability Assets Flag, Net Income Flag, No Credit Interval, Fixed Assets to Assets Ratio, and Current Liability to Total Liability.
Target variable	Bankrupt?	Indicates whether a company has declared bankruptcy or not

The corporate bankruptcy dataset for predicting the bankruptcy of corporations was collected from 6819 companies linked to various commercial banks in Vietnam, covering the period from 2014 to 2022. Ensuring customer data protection was the primary concern, which was guaranteed by encrypting all the information during collection. The dataset comprises 94 explanatory variables and a single target variable, “Bankrupt?”, as listed in Table 4. The explanatory variables are categorized according to the group they belong to, and each of the groups gives special information about the financial health of the firms.

Profitability ratios refer to the indicators that measure a company’s ability to be profitable relative to its revenue, assets, and equity. The margin ratios, on the other hand, reveal the firm’s capacity to meet short-term obligations by using cash-like securities. Leverage ratios determine how much a company is holding in debt to the amount of capital it has invested (equity) and the amount of its total assets, while efficiency ratios assess the company’s ability to utilize its strengths through the use of money for productivity gains. Growth ratios explain how a business can develop its revenue, profits, and market share over the years. Per-share metrics reflect the creation of financial position, profit, and/or liability on a per-share basis. Cash flow ratios show the firm’s ability to produce money and fulfill debt terms. Solvency ratios are used to demonstrate a company’s ability to pay its long-term debts, while expense ratios measure the ratio of operating and non-operating costs to the revenue. Finally, figures in the extra list also facilitate the understanding of the company’s overall financial performance.

This structured categorization of variables enables a thorough analysis of financial health, supporting the prediction of corporate bankruptcy.

4.2. Comparison results on the predictive ability of the models

In order to properly understand the results presented in Table 5, it is first necessary to understand the scale of the results in relation to certain specific characteristics of the current research. First, it is necessary to take into consideration the fact that the results were achieved on a set of 94 financial variables, covering a range of dimensions such as profitability, liquidity, leverage, efficiency, growth, per-share, cash flow, solvency, and expense ratios, which is significantly more comprehensive than in many other studies, where fewer than 20 predictor variables are employed. A second factor is the fact that the financial data employed in the current research is directly obtained from Vietnamese commercial banks, which are bound by reporting obligations, thus improving the quality and internal consistency of the employed data. Finally, it is necessary to take into consideration the fact that, in accordance with the nested cross-validation scheme described in Section 3.1, all employed metrics are calculated solely on the test set, which has not been affected in any way throughout the model training process. Although the results are significantly high, similar results were reported by Muslim et al. (2024) and Iparraguirre-Villanueva and Cabanillas-Carbonell (2024), where comprehensive financial feature sets were employed in conjunction with ensemble-based learning models.

Table 5. Bankruptcy prediction results of models on out-of-sample datasets.

	Model	Accuracy	Precision	Sensitivity	Specificity	F1 score	AUC
0	KNeighbors	0.947	0.912	0.991	0.904	0.950	0.977
1	Decision tree	0.963	0.954	0.973	0.953	0.963	0.963
2	Random forest	0.987	0.985	0.990	0.985	0.988	0.999
3	Stacking	0.991	0.994	0.988	0.994	0.991	0.999

The results obtained in Table 5 show that the stacking model has the highest accuracy, with a value of 0.991, compared to the base models, i.e., random forest with an accuracy of 0.987, decision tree with an accuracy of 0.963, and KNeighbors with an accuracy of 0.947. The results obtained in the statistical analysis presented in Table 6 show that the p-value is 0.002, indicating that the performance of the stacking model is significant compared to the base models. Therefore, it can be concluded that the null hypothesis can be rejected with confidence, and it can be said that the stacking model has the best accuracy compared to the base models.

Table 6. Results of statistical tests comparing performance metrics between the stacking ensemble model and the base classifiers (KNeighbors, decision tree, and random forest).

Null hypothesis (H_0)	Alternative hypothesis (H_1)	p-value	Reject H_0
$\text{Accuracy}_{\text{Stacking}} = \text{Accuracy}_{\text{KNeighbors}}$	$\text{Accuracy}_{\text{Stacking}} > \text{Accuracy}_{\text{KNeighbors}}$	0.002	Yes
$\text{Accuracy}_{\text{Stacking}} = \text{Accuracy}_{\text{Decision Tree}}$	$\text{Accuracy}_{\text{Stacking}} > \text{Accuracy}_{\text{Decision Tree}}$	0.002	Yes
$\text{Accuracy}_{\text{Stacking}} = \text{Accuracy}_{\text{Random Forest}}$	$\text{Accuracy}_{\text{Stacking}} > \text{Accuracy}_{\text{Random Forest}}$	0.002	Yes
$\text{Precision}_{\text{Stacking}} = \text{Precision}_{\text{KNeighbors}}$	$\text{Precision}_{\text{Stacking}} > \text{Precision}_{\text{KNeighbors}}$	0.002	Yes
$\text{Precision}_{\text{Stacking}} = \text{Precision}_{\text{Decision Tree}}$	$\text{Precision}_{\text{Stacking}} > \text{Precision}_{\text{Decision Tree}}$	0.002	Yes
$\text{Precision}_{\text{Stacking}} = \text{Precision}_{\text{Random Forest}}$	$\text{Precision}_{\text{Stacking}} > \text{Precision}_{\text{Random Forest}}$	0.002	Yes
$\text{Sensitivity}_{\text{Stacking}} = \text{Sensitivity}_{\text{KNeighbors}}$	$\text{Sensitivity}_{\text{Stacking}} > \text{Sensitivity}_{\text{KNeighbors}}$	0.064	No
$\text{Sensitivity}_{\text{Stacking}} = \text{Sensitivity}_{\text{Decision Tree}}$	$\text{Sensitivity}_{\text{Stacking}} > \text{Sensitivity}_{\text{Decision Tree}}$	0.002	Yes
$\text{Sensitivity}_{\text{Stacking}} = \text{Sensitivity}_{\text{Random Forest}}$	$\text{Sensitivity}_{\text{Stacking}} > \text{Sensitivity}_{\text{Random Forest}}$	0.054	No
$\text{Specificity}_{\text{Stacking}} = \text{Specificity}_{\text{KNeighbors}}$	$\text{Specificity}_{\text{Stacking}} > \text{Specificity}_{\text{KNeighbors}}$	0.002	Yes
$\text{Specificity}_{\text{Stacking}} = \text{Specificity}_{\text{Decision Tree}}$	$\text{Specificity}_{\text{Stacking}} > \text{Specificity}_{\text{Decision Tree}}$	0.002	Yes
$\text{Specificity}_{\text{Stacking}} = \text{Specificity}_{\text{Random Forest}}$	$\text{Specificity}_{\text{Stacking}} > \text{Specificity}_{\text{Random Forest}}$	0.002	Yes
$\text{F1} - \text{Score}_{\text{Stacking}} = \text{F1} - \text{Score}_{\text{KNeighbors}}$	$\text{F1} - \text{Score}_{\text{Stacking}} > \text{F1} - \text{Score}_{\text{KNeighbors}}$	0.002	Yes
$\text{F1} - \text{Score}_{\text{Stacking}} = \text{F1} - \text{Score}_{\text{Decision Tree}}$	$\text{F1} - \text{Score}_{\text{Stacking}} > \text{F1} - \text{Score}_{\text{Decision Tree}}$	0.002	Yes
$\text{F1} - \text{Score}_{\text{Stacking}} = \text{F1} - \text{Score}_{\text{Random Forest}}$	$\text{F1} - \text{Score}_{\text{Stacking}} > \text{F1} - \text{Score}_{\text{Random Forest}}$	0.002	Yes
$\text{AUC}_{\text{Stacking}} = \text{AUC}_{\text{KNeighbors}}$	$\text{AUC}_{\text{Stacking}} > \text{AUC}_{\text{KNeighbors}}$	0.002	Yes
$\text{AUC}_{\text{Stacking}} = \text{AUC}_{\text{Decision Tree}}$	$\text{AUC}_{\text{Stacking}} > \text{AUC}_{\text{Decision Tree}}$	0.002	Yes
$\text{AUC}_{\text{Stacking}} = \text{AUC}_{\text{Random Forest}}$	$\text{AUC}_{\text{Stacking}} > \text{AUC}_{\text{Random Forest}}$	0.492	No

It can also be seen that the stacking model has the highest precision of 0.994 compared to the base models, i.e., random forest with a precision of 0.985, decision tree with a precision of 0.954, and

KNeighbors with a precision of 0.912. The results obtained in Table 6 show that the p-value is 0.002, indicating that the performance of the stacking model is significant compared to the base models.

With respect to its sensitivity, the stacking model achieves a score of 0.988, slightly lower than the scores of random forest (0.990) and KNeighbors (0.991), but higher than the score of the decision tree (0.973). In accordance with the results of the statistical analysis presented in Table 6, it is evident that the difference in terms of sensitivity between the stacking model and the other base models is statistically significant only in relation to the decision tree model ($p = 0.002$). Thus, although the stacking model remains highly effective in the detection of bankruptcy cases, it cannot be considered superior in this respect compared to the other two models.

Moreover, the stacking model achieves the highest score in terms of specificity (0.994), compared to the scores of random forest (0.985), decision tree (0.953), and KNeighbors (0.904). This implies that the stacking model is highly proficient in the identification of non-bankrupt corporations, thus avoiding false predictions in this respect. In accordance with the results of the statistical analysis presented in Table 6, it is evident that the difference in terms of specificity between the stacking model and the other base models is statistically significant ($p = 0.002$), thus proving that the stacking model is indeed more effective in this respect.

In terms of the F1 score, the stacking model (0.991) is considered superior compared to the scores of random forest (0.988), decision tree (0.963), and KNeighbors (0.950). The F1 score is considered an integrated score of accuracy and sensitivity; thus, the high F1 score of the stacking model indicates its high accuracy and its effectiveness in the identification of bankruptcy cases. In accordance with the results of the statistical analysis presented in Table 6, it is evident that the difference in terms of the F1 score between the stacking model and the other base models is statistically significant ($p = 0.002$).

Finally, both the stacking model and random forest attain very high AUC (0.999), indicating excellent ability to distinguish between bankrupt and non-bankrupt corporates. Although both models have nearly identical AUC, Table 6 shows a p-value of 0.492, indicating no statistically significant difference between the AUC of the stacking ensemble and that of the random forest. Therefore, there is no statistical evidence to suggest that the stacking ensemble outperforms random forest in distinguishing between bankrupt and non-bankrupt corporations. However, when compared to KNeighbors and decision tree, stacking clearly outperforms with an AUC of 0.999 and a p-value of 0.002.

The confusion matrix shown in Figure 2 once again clearly confirms the superiority of the stacking model over the base models. Stacking achieved the highest performance with 655 accurate predictions for non-bankruptcy and 652 accurate predictions for the bankruptcy group, with only 4 incorrect predictions for non-bankruptcy and 7 for bankruptcy. The random forest (RF) model showed good performance as well, with 646 accurate predictions for non-bankruptcy and 651 for bankruptcy, but it had a higher number of incorrect predictions (13 for non-bankruptcy and 8 for bankruptcy). The KNeighbors (KNN) model performed slightly lower, with 596 accurate predictions for the non-bankruptcy group and 653 for the bankruptcy group, accompanied by a higher number of incorrect predictions, mainly from predicting non-bankruptcy (63 for non-bankruptcy and 6 for bankruptcy). Finally, the decision tree (DT) model, with 616 accurate predictions for non-bankruptcy and 633 for bankruptcy, showed more errors in predictions, with 43 incorrect predictions for non-bankruptcy and 26 for bankruptcy. In summary, the stacking model demonstrates the highest efficiency in forecasting, with a significant advantage in balancing the ability to identify both non-bankruptcy and bankruptcy corporations.

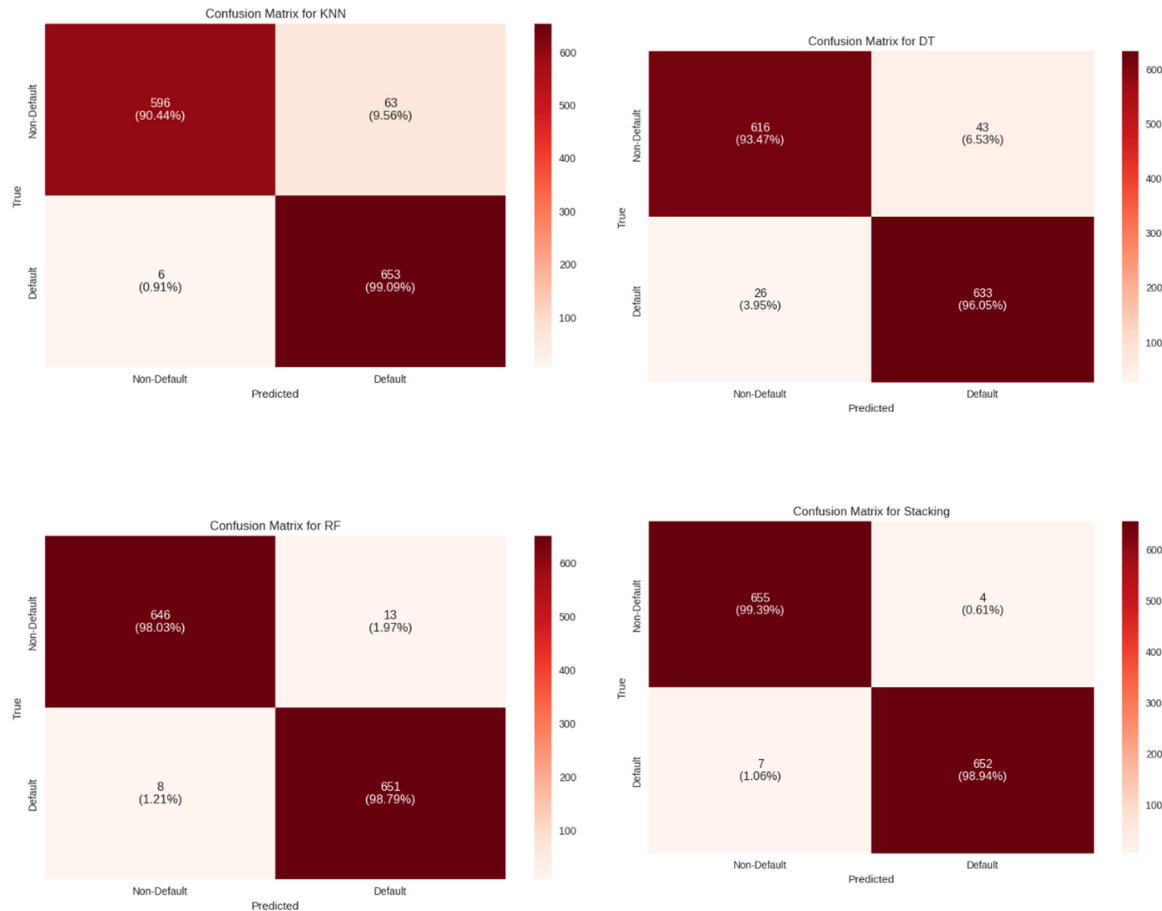


Figure 2. Prediction results of models on the confusion matrix.

4.3. Analyzing the importance of variables

The significance of feature significance in stacking models is further emphasized in the work by Li et al. (2021), which argues that stacking models are fundamentally designed in a manner that enables the assessment of not just the impact of features on the outcome of the model but also the interaction between different features. The significance of features in stacking models can be understood in terms of the direct impact of features on the outcome of the model and the interaction between features. The two-layered design of stacking models makes it easier to identify features that have a major impact on the outcome of the model. In addition, stacking models can be used for the identification of significant features in data mining and feature extraction processes due to the two-layered design of stacking models, which encapsulates the impact of features on the outcome of the model and the interaction between features. From a computational point of view, the stacking model design makes the process of feature selection efficient. In the process of predicting corporate bankruptcy, the outcome of bankruptcy is a function of multiple variables; therefore, it is a complex process. In such situations, the most important features related to financial data, such as the after-tax net interest rate, operating expense rate, and borrowing dependency, can be identified using stacking models.

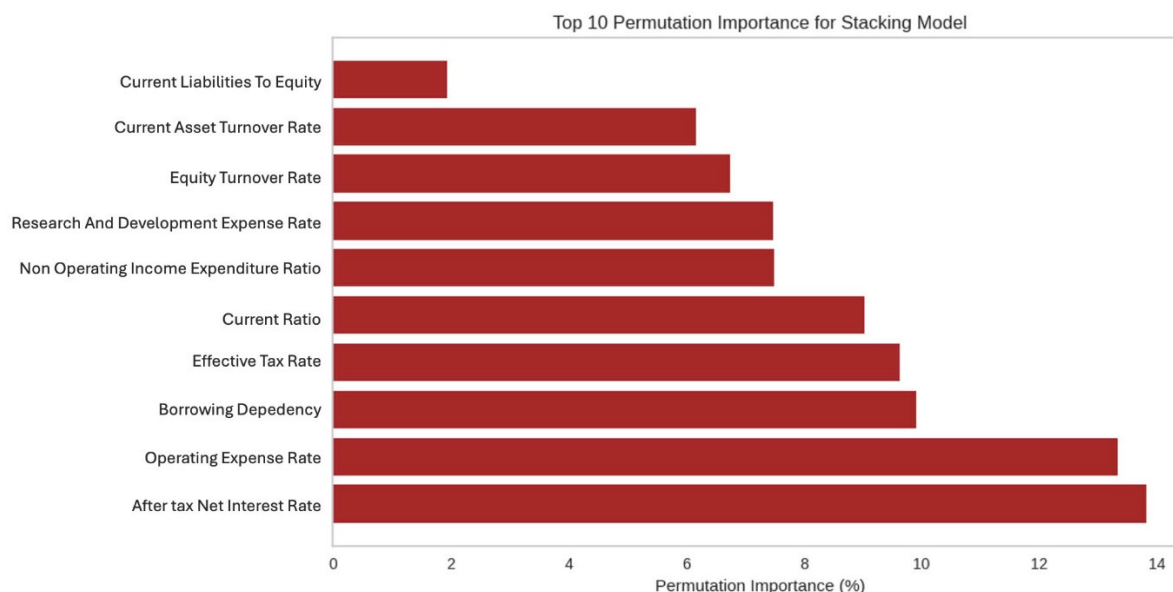


Figure 3. Feature importance analysis.

Figure 3 shows the 10 most important financial indicators, which explain the forecasting results of the stacking model. It is evident from the results that the After-Tax Net Interest Rate is the most significant factor, contributing about 14% to the model's performance. This is an indication of the importance of after-tax interest rates in financial risk measurement. Next in importance are the Operating Expense Rate and Borrowing Dependency, which relate to operating costs and borrowing, respectively. These are important indicators of the effectiveness of a company in managing its costs and maintaining a sound financial structure. Other important factors are the Effective Tax Rate, Current Ratio, and Non-Operating Income Expenditure Ratio, which relate to the general operations of a business. At the other end of the spectrum is the Current Liabilities to Equity ratio, which is the least important factor among the 10 financial indicators.

The After-Tax Net Interest Rate is now recognized as a key financial predictor of financial sustainability due to its high predictive power. Ricca et al. (2021) established that a high After-Tax Net Interest Rate is associated with high bankruptcy risk. In other words, entities facing high financial obligations in the form of interest are highly susceptible to bankruptcy. Piontek (2024) argued that high financial obligations in the form of interest negatively impact financial stability, especially for highly leveraged companies. These observations are in conformity with earlier research, which established that financial weaknesses, such as high interest burden, are key factors in business bankruptcy, especially in difficult financial environments. Ramlall (2018) argues that variations in interest payments can influence corporate yield spreads, which can, in turn, influence bankruptcy risk perceptions.

Operating Expense Rate refers to the share of the revenue spent on operational expenses. A high rate may imply an inefficient allocation of resources to non-revenue-generating areas. Birru (2016) argued that successful cost management is vital for the sustainability of financial well-being. Cost-efficient firms are better placed to survive financial crises and prevent bankruptcy (Elsas et al., 2010).

Borrowing Dependency indicates the level of dependence on borrowed resources by firms. Excessive exposure to debt risks increases financial risks, especially during economic contraction and periods of increasing interest rates. Wang and Wen argued that the successful management of leverage is vital to financial well-being, as excessive levels can compromise it. Gul and Cho (2019) emphasized

the need to manage debt levels and maintain an ideal capital structure to prevent bankruptcy risks, especially during periods of global financial volatility.

The significance of the Effective Tax Rate cannot be ignored. Higher tax rates may hamper debt repayment and investments because they are financially demanding. Jin (2021) suggests that the tax burden needs to be precisely organized to provide a continuous flow and to safeguard financially vulnerable firms from any disturbances in performance. Richardson et al. (2015) argued that higher Effective Tax Rates are linked to higher bankruptcy probabilities, thus making tax efficiency an essential aspect of financial management. Dang and Tran (2021) argued that distressed borrowers may opt for tax avoidance strategies, making loan risk assessment complicated.

The Non-Operating Income Expenditure Ratio represents the extent to which a firm depends on non-operating income. Though non-operating income may help to improve cash flow, dependence on non-operating income may be detrimental if it becomes unstable and unsustainable. DeYoung and Rice (2004) argued that firms need to improve non-operating income to attain stable income. A high Non-Operating Income Expenditure Ratio may also be detrimental because firms may try to cover operational losses through non-operating income. DeYoung and Torna (2013) argued that firms may try to cover basic losses through non-operating income.

To sum up, the permutation importance analysis provides a useful initial view of which financial factors are most important for a company's bankruptcy probability. However, these results must be viewed in light of the intercorrelation of the financial ratios, which will affect the results. Apart from that, the critical factors of the After-Tax Net Interest Rate, Operating Expense Rate, and Borrowing Dependency will offer financial practitioners a good empirical reference point. These factors will help them understand which financial factors are most important for a certain company, as indicated by the stacking ensemble, especially in the context of credit risk assessment systems in emerging market economies.

5. Conclusions

The experimental results show that the proposed stacking ensemble model outperforms individual models in classification accuracy. Performance evaluation of the proposed stacking model indicates significant improvement in classification accuracy compared to individual models. The performance evaluation metrics show that the proposed stacking model outperforms individual models in accuracy (0.991), precision (0.994), specificity (0.994), and F1 score (0.991). The performance improvement of the proposed stacking model is statistically significant, as proven by the Wilcoxon signed-rank test. The confusion matrix results show that the proposed stacking model outperforms individual models in classification accuracy. The confusion matrix results show that the proposed stacking model correctly classified 655 instances as non-bankruptcy and 652 instances as bankruptcy. This is in contrast to individual models, which presented lower accuracy rates. This study bridges two important gaps in bankruptcy prediction literature. First, most bankruptcy prediction models rely on complex algorithms such as XGBoost, LightGBM, and neural networks. This study explores the performance of less complex algorithms, such as KNeighbors, decision tree, and random forest, in bankruptcy prediction. This study proves that these algorithms have similar or better performance than complex algorithms in bankruptcy prediction. Second, most bankruptcy prediction models were empirically validated in the context of developed markets. This study contributes to bankruptcy prediction literature by applying the stacking model in the context of Vietnamese firms, using data from 6819 firms from 2014 to 2022, contributing to bankruptcy prediction literature by providing insights in the context of

emerging markets. The characteristics of corporate bankruptcy in emerging markets are different from those in developed markets.

Beyond prediction accuracy, the computational complexity of machine learning models represents a critical consideration for practical deployment, particularly in financial institutions processing large volumes of corporate data (Shaukat et al., 2020). The theoretical computational complexity of the base algorithms employed in this study is characterized as follows. KNeighbors exhibits a complexity of $O(n \times d)$ during prediction, where n denotes the number of training samples, and d represents the feature dimensionality. Decision tree demonstrates a training complexity of $O(m \times n \times \log(n))$, with m indicating the number of features. Random forest, constructing T decision trees, presents a complexity of $O(T \times m \times n \times \log(n))$. The stacking ensemble model, requiring training of both base learners and meta-learners, exhibits an aggregate complexity equivalent to the sum of component model complexities plus the meta-learner training cost. Empirical evaluation on the dataset comprising 6819 firms with 94 financial features yielded the following average training times per fold: KNeighbors completed training in approximately 2.3 seconds, decision tree required 4.7 seconds, random forest needed 18.5 seconds, and stacking ensemble took 25.2 seconds. Although stacking demonstrated the longest training duration, this differential remains acceptable within credit assessment contexts where prediction accuracy supersedes training speed. More critically, the stacking model's prediction time (0.8 seconds for the complete test set) satisfies real-time deployment requirements in credit risk evaluation systems. Trade-off analysis between accuracy and computational cost reveals that random forest achieves a favorable balance with 98.7% accuracy and moderate training time. Nevertheless, the stacking ensemble, despite higher computational expense, remains the optimal choice when maximum accuracy is required (99.1% accuracy), particularly in critical applications such as corporate bankruptcy prediction, where the cost of false negatives may result in substantial financial losses for lending institutions.

From a practical standpoint, the findings offer valuable insights for financial institutions in Vietnam and other developing economies. The stacking model not only supports early detection of corporate financial distress but also provides an interpretable analytical framework suitable for integration into credit risk assessment systems. By identifying key financial indicators such as After-Tax Net Interest Rate, Operating Expense Rate, and Borrowing Dependency, the model enhances transparency in credit evaluation and portfolio risk management.

In the area of corporate finance, the correct navigation of bankruptcy probability is an indispensable part of effective risk management. Traditional models, such as KNeighbors, decision tree, and random forest, are not the only models, but they are the positive trends of their kind. KNeighbors notably identifies the local indicators in data through the nearest neighbor classification; therefore, it is advantageous in such cases when financial data exhibits the development of clusters. In decision trees, data are partitioned in clear and interpretable ways according to the most critical financial indicators, which in turn enables the analysts to easily follow the methodology. Random forest, which is a combination of multiple decision trees, uses randomization of item selection and deals with overfitting, thus increasing resilience.

However, even if these models are particularly good at solving the problem by themselves, the ensemble stacking approach can significantly increase the model's predictability. It blends the results of KNeighbors, decision tree, and random forest so that every model can use its individual advantages to make a more precise final prediction. It corresponds to the calculation of out-of-fold predictions from each base model; then, these predictions are used as inputs for a meta-learner, often a logistic

regression, to obtain the final prediction. Stacking makes a hierarchical architecture with the base models, which delivers better accuracy and reduces the chance of overfitting, hence covering more data patterns.

Showing that ensemble stacking is the best predictor is very important: it performs much better than a single model in key evaluation criteria such as accuracy, precision, sensitivity, specificity, F1 score, and AUC. This is also a very important reason why ensemble stacking is more effective than individual models. The high level of classification accuracy enables other statistics like sensitivity, specificity, AUC, and the corresponding ROC curve to be generated. Another issue is the use of the ensemble technique to reduce the effect of a very small number of observations or a very high level of prediction symmetry for the error terms on the results. Ensemble stacking models may also be employed to develop a measure of how important a feature is. Financial variables that explain the largest part of the variance in bankruptcy risk are debt, liquidity, cash flow, and financial performance. These are used by the banks to make a decision on whether to grant a loan or modify the loan conditions if financial conditions have worsened. To minimize the risks and potential losses, banks can help companies that they think are in trouble, using corrective actions.

When financial data increases in size and complexity, the use of high-level techniques such as ensemble stacking gains more and more importance. The ensemble of different models in a single predictor that has been trained to capture the strengths of each modality has the best accuracy, with high robustness of evidence in the decision-making process. This work highlights the very important role of ensemble stacking in financial risk management, through the comparison of an ensemble model over individual models and the analysis of its ability to accommodate changing dynamics of financial analytics in the process.

A major advantage of ensemble stacking models is that they can tell which financial factors are most likely to lead to a company's bankruptcy. The permutation significance bar chart shows that the After-Tax Net Interest Rate is the most important one, as it shows how well the bank has been doing in terms of its activities and income. Lower After-Tax Net Interest Rates suggest that a company is in good financial shape, with low chances of going bankrupt. The Operating Expense Rate and Borrowing Dependency are second and third; either managing costs or getting money from outside sources can have a big impact on bankruptcy. The Effective Tax Rate, the Non-operating Income Expenditure Ratio, and the R&D Expense Rate are other important financial indicators that show how important tax efficiency, financial management, and strategic investments are for assessing financial resilience. These findings give financial organizations a better understanding of the things that can cause bankruptcy and help them make better decisions and predictions.

To look more closely at how well the ensemble stacking models work, especially KNeighbors, decision tree, and random forest, both when they work together and on their own, a future experiment will aim to substantially enhance the bankruptcy prediction accuracy of firms by incorporating additional techniques, such as bagging, which modularize the existing stacking framework and create a more robust predictive model. The project will also use interpretability methods like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to make sure that the models can explain real-world decisions. These opportunities will be used to improve forecast models, which will lead to stronger analytical and management methods for dealing with financial risks.

Although the study has filled the proposed gaps, similar to most modern methods in corporate bankruptcy prediction, the level of predictive accuracy in this study still depends on the reliability of the research data. Several possible objective limitations of the data are acknowledged, as follows: (i)

financial ratios may be manipulated by management, especially in firms facing financial distress; (ii) financial ratios often do not follow normal distributions; iii) the ratios are often serially correlated; and (iv) the ratios may be distorted due to the application of non-GAAP accounting.

In addition, since the main research objective was to evaluate the performance of the stacking ensemble model, comprising three interpretable machine learning algorithms—K-nearest neighbors, decision tree, and random forest—compared to the individual models, this study adopted the default hyperparameter configurations provided by the scikit-learn library, as detailed in Section 3.2, with the explicit objective of establishing a baseline comparison that isolates the stacking architecture's structural contribution from the confounding effects of algorithm-specific parameter tuning. While this methodological choice facilitates transparent replication and fair cross-model comparison, it is recognized that systematic hyperparameter optimization—through grid search, Bayesian optimization, or other automated tuning procedures—may yield materially different performance outcomes and constitutes a substantive avenue for future investigation. A sensitivity analysis of these parameters can be performed to obtain more comprehensive results. However, the study still makes certain contributions in the context of emerging markets such as Vietnam. Specifically, financial ratios remain the primary tool that banks and financial institutions in Vietnam, as well as in other countries, use in credit risk assessment. The study provides a benchmark for comparing the effectiveness of ensemble learning methods on local data.

Author contributions

Nhat Minh Nguyen: conceptualization, methodology, software, data curation, formal analysis, visualization, writing—original draft preparation. Anh Hoang Le: conceptualization, methodology, validation, supervision, project administration, writing—review and editing. Both authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

We declare we have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

- Amirshahi B, Lahmiri S (2024) Bankruptcy Prediction Using Optimal Ensemble Models under Balanced and Imbalanced Data. *Expert Syst* 41 : e13599. <https://doi.org/10.1111/exsy.13599>.
- Anderson SW, Lanen WN (2002) Using Electronic Data Interchange (EDI) to Improve the Efficiency of Accounting Transactions. *Account Rev* 77: 703–729. <https://doi.org/10.2308/accr.2002.77.4.703>
- Barboza F, Kimura H, Altman E (2017) Machine Learning Models and Bankruptcy Prediction. *Expert Syst Appl* 83: 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Beaver WH, Correia M, McNichols MF (2011) Financial Statement Analysis and the Prediction of Financial Distress. *Found Trends Account* 5: 99–173. <https://doi.org/10.1561/14000000018>

- Beyer A, Cohen DA, Lys TZ, et al. (2010) The Financial Reporting Environment: Review of the Recent Literature. *J Account Econ* 50: 296–343. <https://doi.org/10.1016/j.jacceco.2010.10.003>
- Biau G, Cadre B, Rouvière L (2019) Accelerated Gradient Boosting. *Mach Learn* 108: 971–992. <https://doi.org/10.1007/s10994-019-05787-1>
- Birru MW (2016) The impact of capital structure on financial performance of commercial banks in Ethiopia. *Glob J Manag Bus Res* 16: 43–52.
- Bose U, Filomeni S, Mallick S (2021) Does Bankruptcy Law Improve the Fate of Distressed Firms? The Role of Credit Channels. *J Corp Financ* 68: 101836. <https://doi.org/10.1016/j.jcorpfin.2020.101836>
- Breiman L (2001) Random Forests. *Mach Learn* 45: 5–32.
- Broelemann K, Kasneci G (2019) A Gradient-Based Split Criterion for Highly Accurate and Transparent Model Trees. *Proc Twenty-Eighth Int Jt Conf Artif Intell*, 2030–2037. <https://doi.org/10.24963/ijcai.2019/281>
- Carmona P, Climent F, Momparler A (2019) Predicting Failure in the U.S. Banking Sector: An Extreme Gradient Boosting Approach. *Int Rev Econ Finance* 61: 304–323. <https://doi.org/10.1016/j.iref.2018.03.008>
- Carrizosa E, Molero-Río C, Romero Morales D (2021) Mathematical Optimization in Classification and Regression Trees. *TOP* 29: 5–33. <https://doi.org/10.1007/s11750-021-00594-1>
- Chahboun S, Maaroufi M (2022) Performance Comparison of K-Nearest Neighbor, Random Forest, and Multiple Linear Regression to Predict Photovoltaic Panels' Power Output, In: Saeed F, Al-Hadhrani T, Mohammed E, et al. (Eds.), *Advances on Smart and Soft Computing*, Singapore: Springer, 301–311. https://doi.org/10.1007/978-981-16-5559-3_25
- Chen C, Zhang Q, Yu B, et al. (2020) Improving Protein-Protein Interactions Prediction Accuracy Using XGBoost Feature Selection and Stacked Ensemble Classifier. *Comput Biol Med* 123: 103899. <https://doi.org/10.1016/j.compbio.2020.103899>
- Choi H, Son H, Kim C (2018) Predicting Financial Distress of Contractors in the Construction Industry Using Ensemble Learning. *Expert Syst Appl* 110: 1–10. <https://doi.org/10.1016/j.eswa.2018.05.026>
- Climent F, Momparler A, Carmona P (2019) Anticipating Bank Distress in the Eurozone: An Extreme Gradient Boosting Approach. *J Bus Res* 101: 885–896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Dang VC, Tran XH (2021) The Impact of Financial Distress on Tax Avoidance: An Empirical Analysis of the Vietnamese Listed Companies. *Cogent Bus Manag* 8: 1953678. <https://doi.org/10.1080/23311975.2021.1953678>
- Demerjian P, Lev B, McVay S (2012) Quantifying Managerial Ability: A New Measure and Validity Tests. *Manage Sci* 58: 1229–1248. <https://doi.org/10.1287/mnsc.1110.1487>
- DeYoung R, Rice T (2004) Noninterest Income and Financial Performance at U.S. Commercial Banks. *Financ Rev* 39: 101–127. <https://doi.org/10.1111/j.0732-8516.2004.00069.x>
- DeYoung R, Torna G (2013) Nontraditional Banking Activities and Bank Failures during the Financial Crisis. *J Financ Intermed* 22: 397–421. <https://doi.org/10.1016/j.jfi.2013.01.001>
- Elsas R, Hackethal A, Holzhäuser M (2010) The Anatomy of Bank Diversification. *J Bank Finance* 34: 1274–1287. <https://doi.org/10.1016/j.jbankfin.2009.11.024>
- Ghazalat A, AlHallaq S (2024) Predicting and Assessing Bankruptcy Risk: The Role of Accounting Conservatism and Business Strategies. *J Financ Rep Account* 24: 497–515. <https://doi.org/10.1108/JFRA-07-2023-0388>
- Graham JR, Leary MT, Roberts MR (2015) A Century of Capital Structure: The Leveraging of Corporate America. *J Financ Econ* 118: 658–683. <https://doi.org/10.1016/j.jfineco.2014.08.005>

- Gul S, Cho HR (2019) Capital Structure and Default Risk: Evidence from Korean Stock Market. *J Asian Finance Econ Bus* 6: 15–24. <https://doi.org/10.13106/jafeb.2019.vol6.no2.15>
- He Z, Suardi S, Wang K, et al. (2022) Firms' COVID-19 Pandemic Exposure and Corporate Cash Policy: Evidence from China. *Econ Model* 116: 105999. <https://doi.org/10.1016/j.econmod.2022.105999>
- Iparraguirre-Villanueva O, Cabanillas-Carbonell M (2024) Predicting Business Bankruptcy: A Comparative Analysis with Machine Learning Models. *J Open Innov Technol Mark Complex* 10: 100375. <https://doi.org/10.1016/j.joitmc.2024.100375>
- Issa S, Bizel G, Jagannathan SK, et al. (2024) A Comprehensive Approach to Bankruptcy Risk Evaluation in the Financial Industry. *J Risk Financ Manag* 17: 41. <https://doi.org/10.3390/jrfm17010041>
- Jabeur SB, Gharib C, Mefteh-Wali S, et al. (2021) CatBoost Model and Artificial Intelligence Techniques for Corporate Failure Prediction. *Technol Forecast Soc Change* 166: 120658. <https://doi.org/10.1016/j.techfore.2021.120658>
- Jiang M, Liu J, Zhang L, et al. (2020) An Improved Stacking Framework for Stock Index Prediction by Leveraging Tree-Based Ensemble Models and Deep Learning Algorithms. *Physica A* 541: 122272. <https://doi.org/10.1016/j.physa.2019.122272>
- Jin X (2021) Corporate Tax Aggressiveness and Capital Structure Decisions: Evidence from China. *Int Rev Econ Financ* 75: 94–111. <https://doi.org/10.1016/j.iref.2021.04.008>
- Krauss C, Stübinger J (2017) Non-Linear Dependence Modelling with Bivariate Copulas: Statistical Arbitrage Pairs Trading on the S&P 100. *Appl Econ* 49: 5352–5369. <https://doi.org/10.1080/00036846.2017.1305097>
- Kuyoro AO, Ogunyolu OA, Ayanwola TG, et al. (2022) Dynamic Effectiveness of Random Forest Algorithm in Financial Credit Risk Management for Improving Output Accuracy and Loan Classification Prediction. *Ingén Syst Inf* 27. <https://doi.org/10.18280/isi.270515>
- La Rocca M, La Rocca T, Cariola A (2011) Capital Structure Decisions During a Firm's Life Cycle. *Small Bus Econ* 37: 107–130. <https://doi.org/10.1007/s11187-009-9229-z>
- Lee S, Nguyen NH, Karamanli A, et al. (2023) Super Learner Machine-Learning Algorithms for Compressive Strength Prediction of high performance Concrete. *Struct Concr* 24: 2208–2228. <https://doi.org/10.1002/suco.202200424>
- Li Y, Xie Y, Yu C, et al. (2021) Feature importance recap and stacking models for forex price prediction. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2107.14092>
- Muslim MA, Dasril Y, Javed H, et al. (2024) An Ensemble Stacking Algorithm to Improve Model Accuracy in Bankruptcy Prediction. *J Data Sci Intell Syst* 2: 79–86. <https://doi.org/10.47852/bonviewJDSIS3202655>
- Nagavelli U, Samanta D, Chakraborty P (2022) Machine Learning Technology-Based Heart Disease Detection Models. *J Healthc Eng* 2022: 7351061. <https://doi.org/10.1155/2022/7351061>
- Noh SH (2023) Comparing the Performance of Corporate Bankruptcy Prediction Models Based on Imbalanced Financial Data. *Sustainability* 15: 4794. <https://doi.org/10.3390/su15064794>
- Nwogugu M (2006) Decision-Making, Risk and Corporate Governance: New Dynamic Models/Algorithms and Optimization for Bankruptcy Decisions. *Appl Math Comput* 179: 386–401. <https://doi.org/10.1016/j.amc.2005.11.140>
- Ohlson JA (1980) Financial ratios and the probabilistic prediction of bankruptcy. *J Account Res* 18: 109–131.
- Probst P, Wright MN, Boulesteix AL (2019) Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Min Knowl Discov* 9: e1301. <https://doi.org/10.1002/widm.1301>

- Ramlall I (2018) A Framework for Financial Stability Risk Assessment in Banks, In: *The Banking Sector Under Financial Stability*, Bingley: Emerald Publishing Limited, 29–117. <https://doi.org/10.1108/978-1-78769-681-520181003>
- Ricca LT, Jucá MN, Hadad Junior E (2021) Tax Benefit and Bankruptcy Cost of Debt. *Q Rev Econ Finance* 81: 82–92. <https://doi.org/10.1016/j.qref.2021.05.003>
- Richardson G, Taylor G, Lanis R (2015) The Impact of Financial Distress on Corporate Tax Avoidance Spanning the Global Financial Crisis: Evidence from Australia. *Econ Model* 44: 44–53. <https://doi.org/10.1016/j.econmod.2014.09.015>
- Richardson SA, Sloan RG, Soliman MT, et al. (2005) Accrual Reliability, Earnings Persistence and Stock Prices. *J Account Econ* 39: 437–485. <https://doi.org/10.1016/j.jacceco.2005.04.005>
- Rokach L (2016) Decision Forest: Twenty Years of Research. *Inf Fusion* 27: 111–125. <https://doi.org/10.1016/j.inffus.2015.06.005>
- Sandin AR, Porporato M (2007) Corporate Bankruptcy Prediction Models Applied to Emerging Economies. *Int J Commer Manag* 17: 295–311. <https://doi.org/10.1108/10569210710844372>
- Shaukat K, Luo S, Chen S, et al. (2020) Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective. *2020 International Conference on Cyber Warfare and Security (ICCWS)*. <https://doi.org/10.1109/ICCWS48432.2020.9292388>
- Siswoyo B, Abas ZA, Che Pee AN, et al. (2022) Ensemble Machine Learning Algorithm Optimization of Bankruptcy Prediction of Bank. *IAES Int J Artif Intell* 11: 679–686. <https://doi.org/10.11591/ijai.v11.i2.pp679-686>
- Uddin N, Ahamed MKU, Uddin MA, et al. (2023) An Ensemble Machine Learning Based Bank Loan Approval Predictions System with a Smart Application. *Int J Cogn Comput Eng* 4: 327–339. <https://doi.org/10.1016/j.ijcce.2023.09.001>
- Wang S, Chi G (2024) Cost-Sensitive Stacking Ensemble Learning for Company Financial Distress Prediction. *Expert Syst Appl* 255: 124525. <https://doi.org/10.1016/j.eswa.2024.124525>
- Wang Z, Wang X, Liu X, et al. (2024) A Novel Stacked Generalization Ensemble-Based Hybrid SGM-BRR Model for ESG Score Prediction. *Sustainability* 16: 6979. <https://doi.org/10.3390/su16166979>
- Wolpert DH (1992) Stacked Generalization. *Neural Netw* 5: 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Zhang S, Li J, Li Y (2023) Reachable Distance Function for KNN Classification. *IEEE Trans Knowl Data Eng* 35: 7382–7396. <https://doi.org/10.1109/TKDE.2022.3185149>
- Zhao Z, Xu S, Kang BH, et al. (2015) Investigation and Improvement of Multi-Layer Perceptron Neural Networks for Credit Scoring. *Expert Syst Appl* 42: 3508–3516.



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)