

Ensemble Machine Learning Models in Financial Distress Prediction: Evidence from China

Yi Ling¹, Pang Paul Wang^{2*}

¹Urban Governance and Design Society Hub, Hong Kong University of Science and Technology, Guangzhou, China

²School of Accounting and Finance, Faculty of Business, The Hong Kong Polytechnic University, Hong Kong, China

Email: yling300@connect.hkust-gz.edu.cn, *paul.wang@polyu.edu.hk

How to cite this paper: Ling, Y. and Wang, P.P. (2024) Ensemble Machine Learning Models in Financial Distress Prediction: Evidence from China. *Journal of Mathematical Finance*, 14, 226-242. <https://doi.org/10.4236/jmf.2024.142013>

Received: February 7, 2024

Accepted: May 25, 2024

Published: May 28, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Corporate distress signals are important for both institutions and banks when evaluating firms' performances. This paper evaluates five different models in predicting the distress for listed companies in China based on 22 dimensions of financial data from 2014 to 2022. The models include three ensemble machine learning models: Adaboost, Bagging, and Random Forest, as well as a single machine learning model Decision Tree, along with a benchmark Logistic Regression. The comparative analysis found Random Forest to be the most promising method with the highest accuracy ratio and lowest Type I and Type II errors. This paper concludes that ensemble learning models could be an easy-to-replicate and highly efficient tool for institutions and banks to evaluate and predict potential distress in firms.

Keywords

Bankruptcy Prediction, Machine Learning, Ensemble Models

1. Introduction

Firms' distress will seriously influence banks, stakeholders, employees, and other relative actors' benefits. While the distress rate may be a measure for economic development, the detection of bankruptcy ensures the efficiency of commercial credit allocation [1]. Therefore, it is important to precisely monitor corporate potential risk of distress to ensure the stability of the financial system.

Compared with econometric methods, artificial intelligence approaches, including machine learning and ensemble learning, are becoming popular. With datasets containing more indicators, it becomes possible to establish more flexi-

ble relationships between indicators and prediction outcomes compared to simple linear models [2]. They also exhibit better performance and could significantly raise the precision in prediction than traditional methods [3]. The merits of machine learning like boosting and random forest are examined by many empirical research [2] [4] and policy evaluation [5] [6] [7] [8]. Among the many machine learning models, ensemble methods are gaining popularity in many prediction topics for their better performances in classification compared with other approaches because they combine various instances into a more accurate prediction [9], thus improving simple supervised learning methods.

Existing bankruptcy research aims to identify a better model with higher predicting precision through multiple approaches, concerning potential predictors from various dimensions [10] [11] [12]. Prediction in the economic area traditionally hinges on the regression of panel data like logistic regression or probit model [13] [14]. Later work improved these models by introducing different extensions such as duration analysis [15], macroeconomic conditions [16], and unobserved heterogeneity [17]. The seminal work of Campbell *et al.* [18] used multiple logit models to predict bankruptcy over different time periods. Studies using newer statistical models followed, such as the Cox intensity model [19], mixed logit model [20], volatility-adjusted distance-to-default measure [21], dynamic capital structure model [22], and so on.

In bankruptcy prediction, various machine learning methods are deployed and their results imply that machine learning models are more accurate, like those in Kim *et al.* [12]. Since studies of bankruptcies using machine learning techniques with sequential data remain rare and bankruptcy results may come from many periods of production [23], Kim *et al.* [12] provided results three benchmarks machine learning ways (logistic model, support vector machine and random forest) using the Compustat North America dataset from 2007 to 2019 but only provide a one-period prediction. They found that random forest performs better than other methods. Sun *et al.* [24] focused on financial distress prediction and found that Adaboost performs better than other models using 2000 to 2008 Chinese-listed companies data, indicating that Adaboost is more suitable for Chinese-listed companies' distress research. Jones *et al.* [25] processed detailed information about why machine learning techniques outstanding traditional econometric methods in bankruptcy prediction, using high-dimensional data (91 predictor variables) in the gradient boosting model, concluding the best and weakest predictors. Moreover, Jones [11] examined 16 classifiers in US corporate bankruptcies, concluding that though simple classifiers (logit) perform better, "new age" methods like Adaboost and random forest are highly recommended for their good interpretability and easy implementation. Barboza *et al.* [26] tested machine learning models based on five variables from the literature Altman [27], and show what future research should focus on: considering the impact of macroeconomic variables, setting up an easy-to-replicate model, which is consistent with [28], who point out that many existing models reflect neither the

panel property of financial statements nor the common influence of macroeconomic conditions on each company. Keasey and Watson [29], Yeh *et al.* [30] conduct models based on financial ratios and non-financial ratios, indicating the role of non-financial ratios in prediction.

This paper mainly aims to predict listing firms' distress through high-dimensional analysis, evaluating corporate responses based on time-series data in China. We extend the existing literature of ensemble machine learning models in bankruptcy prediction and make the following contributions. First, this paper provides a high-dimensional analysis with 22 parameters of financial and non-financial information for a more recent period leading up to the pandemic. Furthermore, we employed variable selection techniques to remove highly correlated variables and performed hyperparameter optimization. This would provide more accurate assessments of the model performances. Second, this paper considers time-series data to provide a more accurate assessment. In China-listed companies, before being labeled with a distress signal ("ST" or "*ST"), some firms might already demonstrate potential distress in previous years' performance. Thus, it's important to include the time effects on the models. Lastly, this paper provides a more comprehensive assessment of ensemble machine learning models by examining various models from different literature.

The rest of this paper is organized as follows. Section 2 describes the methodology and the various adaptive learning models used in this paper. Section 3 presents and discusses our results. Section 4 summarizes our results.

2. Methodology

2.1. Decision Tree

A decision tree is a popular machine-learning algorithm that is used for both classification and regression tasks. It is a tree-like model of decisions and their possible consequences that is built by recursively splitting the data based on the values of the input features.

At the root of the tree, the entire dataset is considered and the input feature that best separates the data into the different classes or the output variable is chosen as the first split. The data is then split into smaller subsets based on the chosen feature value, and the process is repeated recursively for each subset until a stopping criterion is met, such as reaching a maximum tree depth, achieving a minimum number of data points in a leaf node, or a minimum reduction in impurity.

Each internal node of the decision tree represents a decision based on the value of an input feature, while each leaf node represents a predicted output value or class. The decision rules learned by a decision tree can be easily visualized and interpreted, making it a popular choice for applications where interpretability is important.

Decision trees can be used with different impurity measures to decide which feature to split on at each node. Two common impurity measures are Gini im-

purity and entropy. Gini impurity measures the probability of misclassifying a randomly chosen example from a given class, while entropy measures the amount of disorder or uncertainty in the data. In this paper, the CART algorithm is deployed in a single machine learning model and also set as weak classification in ensemble models, which depends on the Gini index to split the classification [31] [32]. Gini coefficient could be written as:

$$Gini(X_k) = \sum_{k=1}^n P(x_k)(1 - p(x_k)) \quad (1)$$

where P is the probability of choosing classifier k with value X .

Decision trees can also be prone to overfitting, where the tree becomes too complex and captures noise in the data rather than the underlying patterns. To address this issue, techniques such as pruning, ensemble methods like random forests, and early stopping can be used.

Decision trees have been successfully applied in many domains, including medicine, finance, and marketing. They are especially useful in applications where the decision-making process is based on a hierarchical set of rules or features, and where interpretability and transparency are important.

2.2. Bagging Model

Bagging, which stands for Bootstrap Aggregating, is a model of adaptive learning that is commonly used in machine learning [26] [31] [33]. The basic idea behind Bagging is to create multiple models using different subsets of the training data and then combine their predictions to improve accuracy and reduce overfitting.

Figure 1 illustrates how it works in this paper.

In Bagging, the training data are randomly sampled with replacement to create multiple subsets, which are used to train individual models. These models can be trained using any algorithm, such as decision trees, neural networks, or support vector machines. The final prediction is then made by combining the predictions of all the individual models, either by taking a majority vote (for classification problems) or by averaging the predictions (for regression problems).

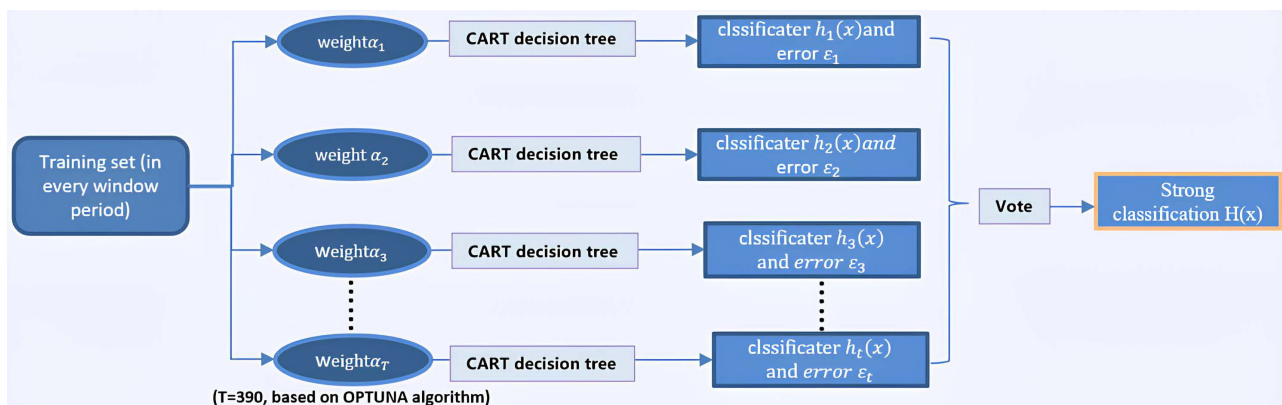


Figure 1. Framework of bagging model.

The use of multiple models trained on different subsets of the data helps to reduce the impact of outliers and noise in the data, as well as to capture different aspects of the underlying patterns. Random sampling also helps to reduce overfitting, as the individual models are more likely to be diverse and less likely to memorize the training data.

Bagging is effective in many real-world applications, such as image and speech recognition, natural language processing, and financial forecasting. However, it may not be suitable for all types of problems, especially those with a small number of training examples or those that require a high degree of interpretability.

2.3. Random Forest

Random Forest is a popular extension of the Bagging algorithm that is also frequently used for classification and regression problems [12] [26] [30]. While Bagging employs all of the features in the subsamples, Random forest only selects a part of the subsamples to train. It was first introduced by Breiman [34] and is based on the concept of decision trees.

In this paper, the Random Forest algorithm works by building a large number of classification and regression trees (CART) on randomly selected subsets of the training data. In this paper, Optuna is deployed and the best number of trees is 133. Each tree is trained on a different subset of the data and with a random subset of the features. This helps to reduce overfitting and improve the generalization of the model. The final prediction is then the average or majority vote of the predictions of all the trees.

One of the main advantages of Random Forest is its ability to handle high-dimensional data and noisy or missing data. It is also less prone to overfitting compared to single decision trees. Moreover, Random Forest can provide feature importance scores, which can be useful for feature selection and understanding the importance of different features in the model.

However, Random Forest also has some limitations, such as its tendency to be computationally expensive, especially when dealing with a large number of trees. Additionally, it can be difficult to interpret the model, especially when dealing with a large number of trees.

2.4. Adaboost Model

Adaptive learning algorithms have been gaining popularity in recent years due to their ability to improve the accuracy of machine learning models [3] [24] [31] [32]. One such algorithm is AdaBoost, which stands for Adaptive Boosting. Introduced in Freund and Schapire [35], AdaBoost is a powerful ensemble learning technique that combines multiple weak classifiers to produce a highly accurate final model.

The AdaBoost algorithm works by iteratively training a sequence of weak classifiers on the training data, and then weighting the data based on the accuracy of the previous classifiers. In this paper weak classifier is the CART decision tree model. The weights assigned to each observation are used to adjust the im-

portance of that observation in subsequent rounds of training. The final model is a weighted combination of the weak classifiers, and its accuracy is typically better than that of any individual weak classifier. **Figure 2** illustrates the framework of Adaboost, when ϵ is determined by the ratio of wrong classifications samples to total samples.

AdaBoost has several advantages over other machine learning techniques, such as its ability to handle high-dimensional data and class imbalance. It is also relatively easy to implement and can be used with a wide range of classification algorithms. However, AdaBoost is not without its limitations, such as its sensitivity to noisy data and tendency to overfit the training data.

2.5. Logistic Model

Logistic regression is a popular machine learning algorithm used for binary classification problems, where the goal is to predict a binary output variable based on one or more input variables Kim *et al.* [12], Bajari *et al.* [33], Beutel *et al.* [10]. It is a type of generalized linear model that uses a logistic function to estimate the probability of the output variable.

In logistic regression, the input features are first linearly combined using a set of weights and biases. The resulting score is then passed through a logistic function, also known as a sigmoid function, which maps the score to a value between 0 and 1. This value represents the estimated probability of the output variable taking a positive class.

The logistic function has an S-shaped curve that starts close to 0 and gradually rises to 1 as the input score increases. The steepness of the curve can be adjusted by changing the slope parameter, which affects the rate at which the probability changes concerning the input score.

The logistic regression model is trained by minimizing a cost function, typically the cross-entropy loss, which measures the difference between the predicted probabilities and the actual labels. This is done using an optimization algorithm such as gradient descent, which updates the weights and biases in the model to minimize the cost function.

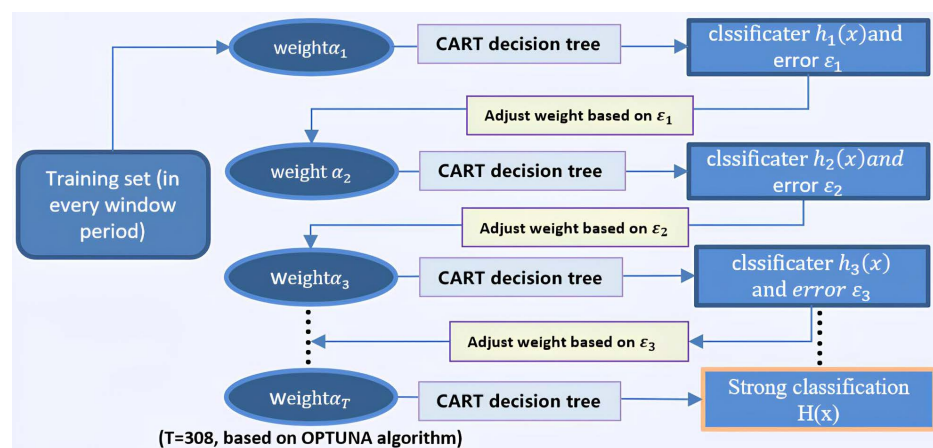


Figure 2. Framework of Adaboost model.

Logistic regression has several advantages, including its simplicity, interpretability, and ability to handle both continuous and categorical input variables. It is also computationally efficient and can be easily extended to handle multi-class classification problems using techniques such as one-vs-all or softmax regression.

Logistic regression has been widely used in many applications, including medical diagnosis, credit risk assessment, and marketing. However, it may not be suitable for all types of problems, especially those with complex decision boundaries or imbalanced classes.

2.6. Optuna in Machine Learning

Optuna is a powerful Python library for hyperparameter optimization, which is a critical component of machine learning [36]. Hyperparameters are the configuration variables that define the behavior and performance of a machine learning model, such as the learning rate, regularization strength, and network architecture. Hyperparameter optimization is the process of finding the optimal combination of hyperparameters that yields the best performance of the model on a given task.

Optuna uses a Bayesian optimization algorithm to efficiently search the hyperparameter space and find the optimal combination of hyperparameters. The algorithm works by building a probabilistic model of the objective function, which maps the hyperparameters to the performance metric of the model, such as accuracy or mean squared error. The model is then used to predict the performance of different hyperparameter settings and guide the search towards promising regions of the hyperparameter space.

Optuna supports various types of hyperparameters, including continuous, discrete, and categorical variables, and provides a flexible and intuitive API for defining the search space and constraints. It also supports parallel and distributed optimization, which allows multiple trials to be run simultaneously on different machines and provides a dashboard for monitoring the progress and results of the optimization. In this paper, Optuna will be deployed in every model before training.

2.7. Model Evaluation

This paper uses five different methods to measure the performance of each model. The first is AUC (accuracy ratio), which is widely deployed as the best criterion to evaluate the performance of the model [31] because it could better depict the performance in imbalanced data classification [12] while demonstrating how the model can separate distressed and non-distress firms. AUC is calculated as the area below ROC (receiver operating characteristic curve).

The second criteria commonly used are Type I and Type II errors, which mainly consider the likelihood of identifying distressed firms and non-distress firms [26]. Type I error is the percentage of firms identified as non-distressed when they are distressed, also known as false positives. Type II error is the per-

centage of firms identified as distressed when they are non-distressed, also known as false negative. Since investors and banks are more concerned with detecting distressed firms, minimizing Type II errors has a higher priority. We denote TP as true positive, the probability of non-distress firms predicted successfully and set it to 0. We denote TN as true negative, representing the probability of distressed firms correctly identified, and labeled it as 1. The variables are determined as follows:

$$\text{Type I Error} = \frac{FP}{FP + TN} \quad (2)$$

$$\text{Type II Error} = \frac{FN}{FN + TP} \quad (3)$$

where FP denotes false positive, and FN represents false negative. Since the confusion matrix records the classification results, this paper provides Type I and II errors through confusion matrices.

In addition to the above two criteria, we also use Precision, Recall, and F1 scores to measure the performance of each model. Precision looks at the number of truly distressed firms from the firms that are predicted as distressed, while Recall seeks the number of truly distressed firms that are successfully detected by the model. The F1 score is the average of Precision and Recall.

3. Empirical Strategy

3.1. Data Description

Chinese mainland listed companies with distressed performance will be labeled as “ST” (special treated), mainly due to consecutive two years’ negative net profit or net capital per share being lower than the face value per share [24], and “*ST” implies the firm is in risk of listing suspension. WIND database provides detailed information on the time for listing changes, and samples during the periods start from “*ST”, “From *ST to ST”, “ST” and “Stock Listing Suspended” are determined as distressed firms. Moreover, since the distress is mainly due to inappropriate business management lasting for several years, WIND database from 2014 to 2022 will be deployed to include the effect of time in prediction.

Utilizing WIND data, 3355 listed companies are selected. In each year, the sample with too many non-values will be excluded from the modeling, then the remaining descriptive data are presented in **Table 1**. Setting every four years as a window period to train, and testing the fifth year based on four-year training. In other words, in each model, data from 2014 to 2017, 2015 to 2018, 2016 to 2019, and 2017 to 2020 will be set into training separately, while data from 2018 to 2021 will be evaluated based on four training respectively. Therefore, data from 2018 to 2021 are subjected to assessment based on the previous four-year training respectively, and the prediction accuracy is determined by the testing outcome from 2018 to 2021. In other words, aggregate accuracy could depict the performance of four times training, and include the information from 2014 to 2021.

Table 1. Summary statistics.

Row Labels	2014	2015	2016	2017	2018	2019	2020	2021	2022	Grand Total
Non-distressed firms	1049	1260	1437	1702	1919	2061	2179	2447	983	15,037
Distressed firms	175	179	181	167	160	163	157	101	16	1299
Share of distressed firms	14.30%	12.44%	11.19%	8.94%	7.70%	7.33%	6.72%	3.96%	1.60%	7.95%
Grand Total	1224	1439	1618	1869	2079	2224	2336	2548	999	16,336

3.2. Variables Selection

Figure 3(a) and **Figure 3(b)** demonstrate the 22 variables selected from WIND based on existing literature [24] [26] [29] [30] [31] [32]. Financial indicators capture and predict the enterprises' financial performance, including the ability to generate profits, Solvency and so on. Non financial ratios evaluate the potential development of the corporate. There are 20 variables covering financial ratios (including profitability, solvency, etc.), one variable for non-financial ratios (*i.e.* whether the audit is performed by big 4 audit firms) [30], and intellectual capital (*i.e.* expenditure in R&D). **Table 2** illustrates the definitions of the variables chosen. Highly correlated variables are excluded from the model.

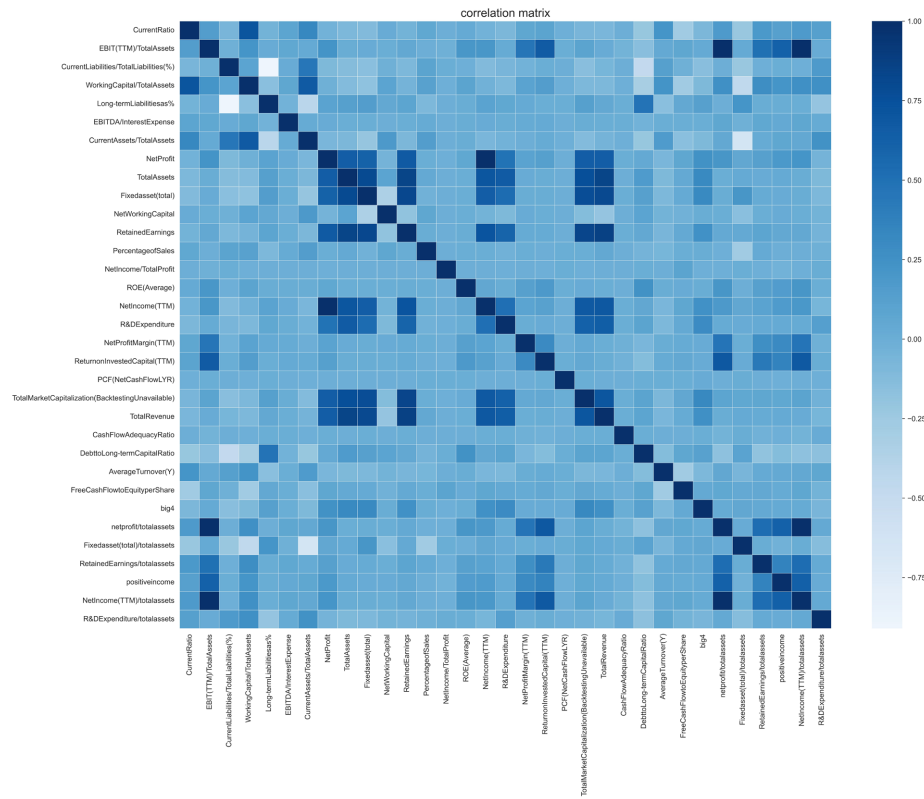
The data in 2022 include too much non-value information, and after deleting those non-value firms, only 16 distressed firms remain, while there are 161 distressed firms. Hence, 2022 data will not be utilized for testing. Distress observations are highly imbalanced and summing up all the observations together, distressed firms only account for 8 percent, with 15,307 overall from 2014 to 2022. Therefore, the synthetic minority over-sampling technique (SMOTE) is deployed in the data, before setting the hyperparameters and training. This is commonly used in many bankruptcy prediction analyses to balance the data [12] [37]. After deploying SMOTE, distressed firms are oversampling from 1299 to 1503, the ratio of distress firms and non-distress firms is 8:10.

3.3. Empirical Results

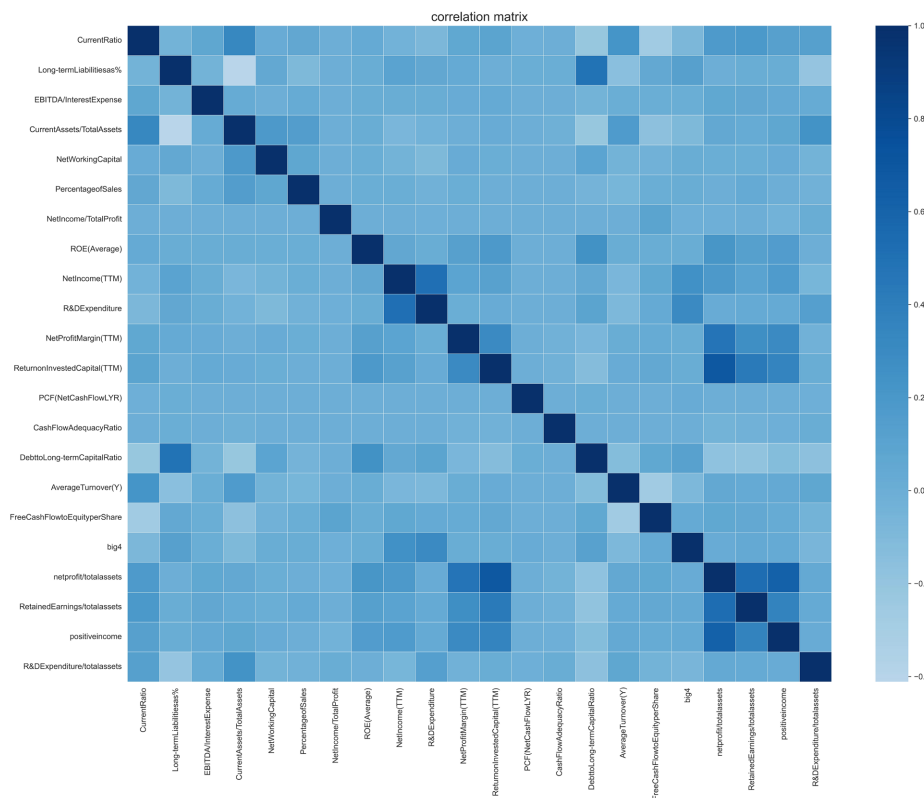
The results are the outcome performance combined with the prediction from 2018 to 2021, based on four different four-year loops of training. This paper evaluates the performance of the five models based on the prediction outcome.

Figure 4 illustrates the comparison of ROC curve, and higher ROC represents better performance. ROC of the Random Forest is 0.94, with 0.01 higher than the Adaboost's ROC curve, and 0.03 higher than Bagging's. All of the ensemble learning models have higher than 0.90 performance, with Random Forest classification becoming the best classification model among the five methodologies, and the decision tree holds the lowest ROC. In addition, Logistic regression performs better than the Decision tree model, with ROC 0.08 higher than 0.74.

Figure 5 further presents the confusion matrix based on the prediction outcome. For the *TP* (true positive) value, Random Forest holds the highest *TP* value than other methodologies, detecting 2192 non-distressed firms in the test set,



(a)



(b)

Figure 3. Correlation matrix. (a) Selected variables; (b) Selected variables excluded variables with high correlations.

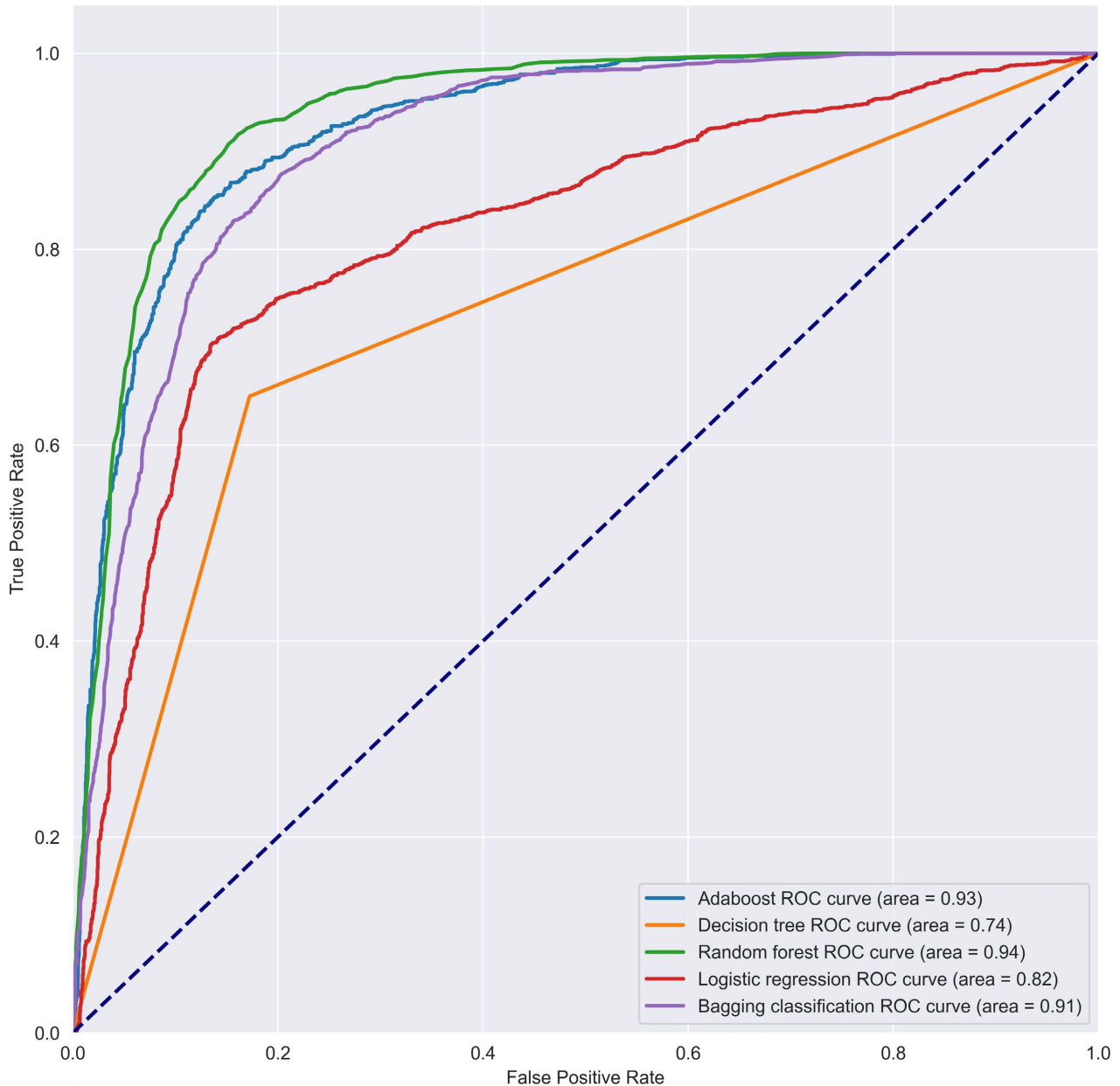


Figure 4. ROC curve.

while Adaboost holds the lowest TP value. However, for TN (true negative) values, Adaboost performs the best and detects 1733 distressed firms, which is 71 units higher than Random Forest, and 538 units higher than logistic regression.

Based on confusion matrices, Table 3 further evaluates the performance of different models. The Precision of Random forest and Bagging are 86.7% and 84%, respectively, which is higher than other models. The Precision of logistic regression is also higher than 80%. The decision tree holds the lowest Precision. In addition, when comparing the recall score among ensemble models, Adaboost holds the highest recall score (88.6%), 3.7% higher than random forest and 12.7% higher than bagging. Different from precision, the recall score of logistic

Table 2. Variables definitions.

Dimensions	Definition
Financial Indicators	Current Ratio
	Long-Term Liabilities
	EBITDA/Interest Expense
	Current Assets/Total Assets
	Net Working Capital
	Percentage of Sales
	Net Income/Total Profit
	ROE (Average)
	Net Income (TTM)
	R & D Expenditure
	Net Profit Margin (TTM)
	Return on Invested Capital (TTM)
	PCF (Net Cash Flow LYR)
	Cash Flow Adequacy Ratio
	Debt to Long-term Capital Ratio
	Average Turnover (Y)
Free Cash Flow to Equity per Share	
Net Profit/Total Assets	
Retained Earnings/Total Assets	
Positive Income*	
Intellectual Capital	R & D Expenditure/Total Assets
Non Financial Indicators	Big Four**

Note: *1 if net income > 0; 0 otherwise. ***1 if audit firm is PwC, Deloitte, E & Y, or KPMG; 0 otherwise.

Table 3. Classification performance.

	Precision	Recall	F1 score	Type I	Type II
Adaboost Classification	79.2%	88.6%	83.6%	11.4%	18.6%
Random Forest Classification	86.7%	84.9%	85.8%	10.4%	15.1%
Bagging Classification	84.0%	75.9%	79.8%	11.5%	24.1%
Decision Tree Classification	75.1%	65.0%	69.7%	17.2%	35.0%
Logistic Regression	82.3%	61.1%	70.1%	10.5%	38.9%

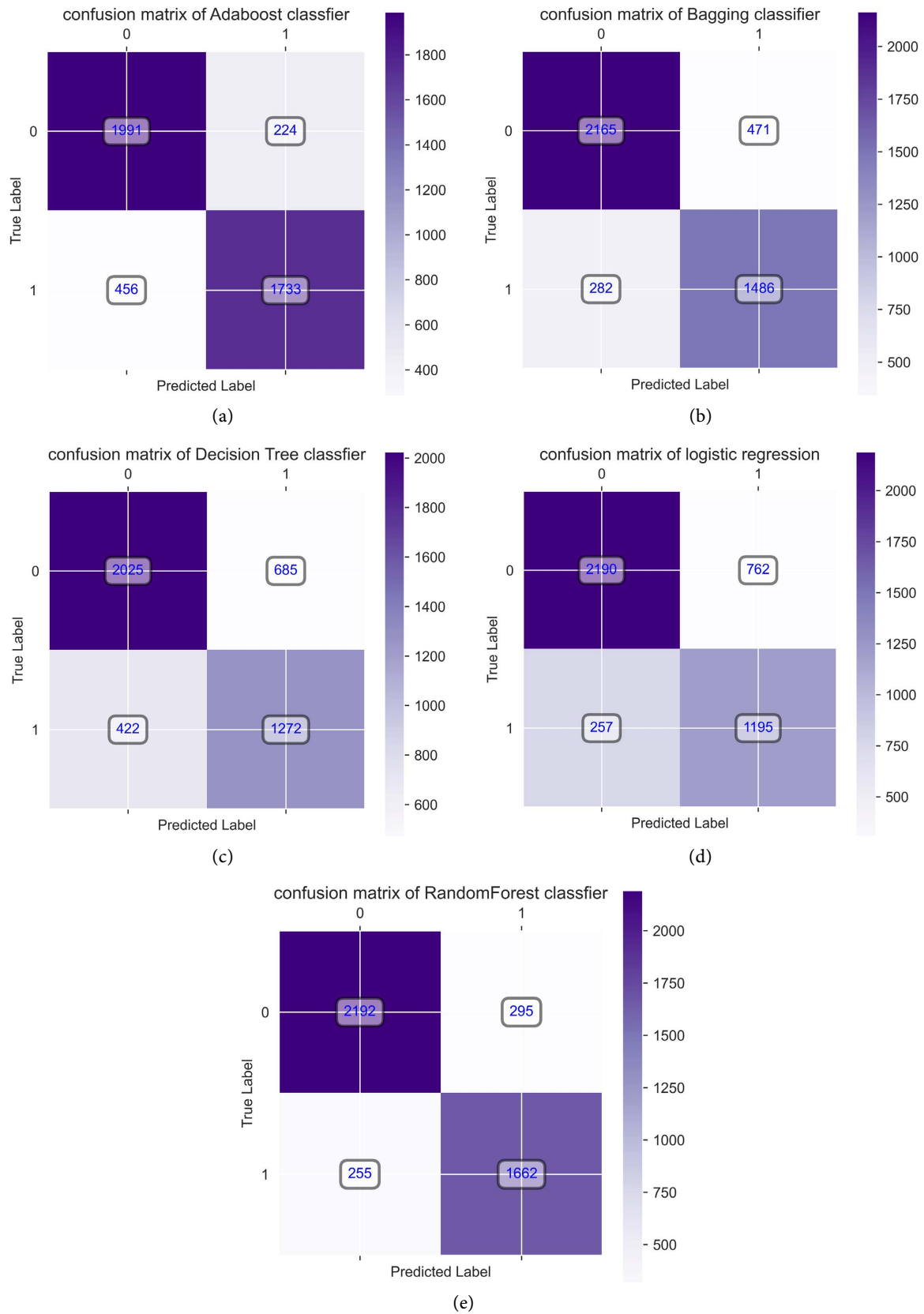


Figure 5. Confusion matrix. (a) Confusion matrix: Adaboost; (b) Confusion matrix: Random forest; (c) Confusion matrix: Bagging; (d) Confusion matrix: Decision tree; (e) Confusion matrix: Logistic regression.

regression is only 61.1% and the lowest. F1 score averages the precision and recall score, and random forest performs best among all the models, with 85.5% in F1 score, with 2.2% higher than Adaboost and 6% higher than Bagging models. Logistic regression and decision tree share close F1 score, with logistic regression 0.4% higher than the decision tree. Above all, combining precision and recall score, ensemble models are better than other methodologies.

The second important evaluation is Type I and Type II errors, with lower errors equating to better performance. Random forest shows the lowest in both Type I and Type II errors among ensemble models. Decision tree as a single machine learning model holds the highest Type I error with 17.2%, while logistic regression holds the highest Type II error with 38.9%, which is nearly 20% higher than Random Forest. Type I error for Random Forest is 10.4%, followed by Adaboost (11.4%), then Bagging (11.5%). Type II error for Random Forest is 15.1%, followed by Adaboost (18.6%), then Bagging (24.5%). Since Type II error is more important in the topic of distress detection, logistic regression is the least recommended model in detecting distressed firms. Additionally, Random Forest is a promising model that's slightly better than other ensemble models.

Combining the two criteria together, though it is hard to rank the performance of the five models, since the training samples SMOTE selects every time are different, random forest performs slightly better and is more suitable for detecting distress than other ensemble learning models like Boosting and Bagging. Moreover, ensemble learning models perform more promising than single machine learning models and traditional models.

4. Conclusions

This study provides easy-to-replicate models that utilize the latest dataset with 22 dimensions after variable selection techniques, hyperparameter optimization, and time effect by setting window periods through the classification. The results contribute to the existing literature that ensemble models perform better than single machine learning models and traditional econometric models. This paper conducted a comparative analysis based on ensemble learning models Random forest, Bagging, Adaboost, single machine learning model Decision Tree, and econometrical model Logistical Regression. All are designed in four window periods for training. Random forest model is found to provide a more efficient result based on the data and variables chosen, which means that Random forest could be a useful tool for clients when making investment decisions. This could supplement banks and audit firms when evaluating the performance of firms.

Our study has several limitations due to limited data availability and computational resources. We only utilized information from listed firms, but investors and banks often lend to non-listed firms. Data on private firms will make the analysis more comprehensive. Our models make single-period predictions and, therefore, cannot provide survival probabilities over time. In addition, the shocks from industry might significantly impact bankruptcy [38], this work did

not consider any industry effects on firm bankruptcy. Thus, future work extensions can include multi-period models and information on private firms.

Funding

This work was supported by Start-Up Funding (ID: A0043272) of Pang Paul Wang from the Hong Kong Polytechnic University.

Conflicts of Interest

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yi Ling and Pang Paul Wang. The first draft of the manuscript was written by Yi Ling and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

- [1] Smiti, S. and Soui, M. (2020) Bankruptcy Prediction Using Deep Learning Approach Based on Borderline Smote. *Information Systems Frontiers*, **22**, 1067-1083. <https://doi.org/10.1007/s10796-020-10031-6>
- [2] Varian, H.R. (2014) Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, **28**, 3-28. <https://doi.org/10.1257/jep.28.2.3>
- [3] Wang, G. and Ma, J. (2011) Study of Corporate Credit Risk Prediction Based on Integrating Boosting and Random Subspace. *Expert Systems with Applications*, **38**, 13871-13878. <https://doi.org/10.1016/j.eswa.2011.04.191>
- [4] Chen, J.M. (2021) An Introduction to Machine Learning for Panel Data. *International Advances in Economic Research*, **27**, 1-16. <https://doi.org/10.1007/s11294-021-09815-6>
- [5] Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A. and Salvestrini, V. (2018) Targeting with Machine Learning: An Application to a Tax Rebate Program in Italy. *Journal of Economic Behavior & Organization*, **156**, 86-102. <https://doi.org/10.1016/j.jebo.2018.09.010>
- [6] Athey, S. (2018) The Impact of Machine Learning on Economics. In: Agrawal, A., Gans, J. and Goldfarb, A., Eds., *The Economics of Artificial Intelligence. An Agenda*, University of Chicago Press, Chicago, 507-547. <https://doi.org/10.7208/chicago/9780226613475.003.0021>
- [7] Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z. (2015) Prediction Policy Problems. *American Economic Review*, **105**, 491-495. <https://doi.org/10.1257/aer.p.20151023>
- [8] Mullainathan, S. and Spiess, J. (2017) Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, **31**, 87-106. <https://doi.org/10.1257/jep.31.2.87>
- [9] Mayr, A., Binder, H., Gefeller, O. and Schmid, M. (2014) The Evolution of Boosting Algorithms. *Methods of Information in Medicine*, **53**, 419-427. <https://doi.org/10.3414/ME13-01-0122>

- [10] Beutel, J., List, S. and von Schweinitz, G. (2019) Does Machine Learning Help Us Predict Banking Crises? *Journal of Financial Stability*, **45**, Article ID: 100693. <https://doi.org/10.1016/j.jfs.2019.100693>
- [11] Jones, S. (2017) Corporate Bankruptcy Prediction: A High Dimensional Analysis. *Review of Accounting Studies*, **22**, 1366-1422. <https://doi.org/10.1007/s11142-017-9407-1>
- [12] Kim, H., Cho, H. and Ryu, D. (2022) Corporate Bankruptcy Prediction Using Machine Learning Methodologies with a Focus on Sequential Data. *Computational Economics*, **59**, 1231-1249. <https://doi.org/10.1007/s10614-021-10126-5>
- [13] Ohlson, J.A. (1980) Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, **18**, 109-131. <https://doi.org/10.2307/2490395>
- [14] Zmijewski, M.E. (1984) Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, **22**, 59-82. <https://doi.org/10.2307/2490859>
- [15] Shumway, T. (2001) Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, **74**, 5-32. <https://doi.org/10.1086/209665>
- [16] Bonfim, D. (2009) Credit Risk Drivers: Evaluating the Contribution of Firm Level Information and of Macroeconomic Dynamics. *Journal of Banking and Finance*, **33**, 281-299. <https://doi.org/10.1016/j.jbankfin.2008.08.006>
- [17] Dakovic, R., Czado, C. and Berg, D. (2010) Bankruptcy Prediction in Norway: A Comparison Study. *Applied Economics Letters*, **17**, 1739-1746. <https://doi.org/10.1080/13504850903299594>
- [18] Campbell, J.Y., Hilscher, J. and Szilagyi, J. (2008) In Search of Distress Risk. *Journal of Finance*, **63**, 2899-2939. <https://doi.org/10.1111/j.1540-6261.2008.01416.x>
- [19] Figlewski, S., Frydman, H. and Liang, W. (2012) Modeling the Effect of Macroeconomic Factors on Corporate Default and Credit Rating Transitions. *International Re-view of Economics and Finance*, **21**, 87-105. <https://doi.org/10.1016/j.iref.2011.05.004>
- [20] Kukuk, M. and Rönning, M. (2013) Corporate Credit Default Models: A Mixed Logit Approach. *Review of Quantitative Finance and Accounting*, **40**, 467-483. <https://doi.org/10.1007/s11156-012-0281-4>
- [21] Jessen, C. and Lando, D. (2015) Robustness of Distance-to-Default. *Journal of Banking and Finance*, **50**, 493-505. <https://doi.org/10.1016/j.jbankfin.2014.05.016>
- [22] Glover, B. (2016) The Expected Cost of Default. *Journal of Financial Economics*, **119**, 284-299. <https://doi.org/10.1016/j.jfineco.2015.09.007>
- [23] Kim, H., Cho, H. and Ryu, D. (2020) Corporate Default Predictions Using Machine Learning: Literature Review. *Sustainability*, **12**, Article No. 6325. <https://doi.org/10.3390/su12166325>
- [24] Sun, J., Jia, M.-Y. and Li, H. (2011) AdaBoost Ensemble for Financial Distress Prediction: An Empirical Comparison with Data from Chinese Listed Companies. *Expert Systems with Applications*, **38**, 9305-9312. <https://doi.org/10.1016/j.eswa.2011.01.042>
- [25] Jones, S., Johnstone, D. and Wilson, R. (2017) Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. *Journal of Business Finance & Accounting*, **44**, 3-34. <https://doi.org/10.1111/jbfa.12218>
- [26] Barboza, F., Kimura, H. and Altman, E. (2017) Machine Learning Models and Bankruptcy Prediction. *Expert Systems with Applications*, **83**, 405-417. <https://doi.org/10.1016/j.eswa.2017.04.006>

- [27] Altman, E.I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, **23**, 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- [28] Nam, C.W., Kim, T.S., Park, N.J. and Lee, H.K. (2008) Bankruptcy Prediction Using a Discrete-Time Duration Model Incorporating Temporal and Macroeconomic Dependencies. *Journal of Forecasting*, **27**, 493-506. <https://doi.org/10.1002/for.985>
- [29] Keasey, K. and Watson, R. (1987) Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti's Hypotheses. *Journal of Business Finance and Accounting*, **14**, 335-354. <https://doi.org/10.1111/j.1468-5957.1987.tb00099.x>
- [30] Yeh, C.-C., Chi, D.-J. and Lin, Y.-R. (2014) Going-Concern Prediction Using Hybrid Random Forests and Rough Set Approach. *Information Sciences*, **254**, 98-110. <https://doi.org/10.1016/j.ins.2013.07.011>
- [31] Du Jardin, P. (2016) A Two-Stage Classification Technique for Bankruptcy Prediction. *European Journal of Operational Research*, **254**, 236-252. <https://doi.org/10.1016/j.ejor.2016.03.008>
- [32] Karas and Režňáková, M. (2017) The Stability of Bankruptcy Predictors in the Construction and Manufacturing Industries at Various Times before Bankruptcy. *E+M Ekonomie a Management*, **20**, 116-133. <https://doi.org/10.15240/tul/001/2017-2-009>
- [33] Bajari, P., Nekipelov, D., Ryan, S.P. and Yang, M. (2015) Machine Learning Methods for Demand Estimation. *American Economic Review*, **105**, 481-485. <https://doi.org/10.1257/aer.p20151021>
- [34] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [35] Freund, Y. and Schapire, R.E. (1997) A Decision-Theoretic Generalization of Online Learning and an Application to Boosting. *Journal of Computer and System Sciences*, **55**, 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- [36] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019) Optuna: A Next-Generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, 4-8 August 2019, 2623-2631. <https://doi.org/10.1145/3292500.3330701>
- [37] Zhou, L. (2013) Performance of Corporate Bankruptcy Prediction Models on Imbalanced Dataset: The Effect of Sampling Methods. *Knowledge-Based Systems*, **41**, 16-25. <https://doi.org/10.1016/j.knosys.2012.12.007>
- [38] Chava, S. and Jarrow, R.A. (2004) Bankruptcy Prediction with Industry Effects. *Review of Finance*, **8**, 537-569. <https://doi.org/10.1093/rof/8.4.537>