Received: 11 January 2023 | Revised: 8 March 2023 | Accepted: 14 March 2023 | Published online: 16 March 2023

Journal of Data Science and Intelligent Systems
2024, Vol. 2(2) 79–86
DOI: 10.47852/bonviewJDSIS3202655

RESEARCH ARTICLE

An Ensemble Stacking Algorithm to Improve Model Accuracy in Bankruptcy Prediction

Much Aziz Muslim[1,*], Yosza Dasril[1], Haseeb Javed[2], Alamsyah[3], Jumanto[3], Wiena Faqih Abror[3], Dwika Ananda Agustina Pertiwi[3] and Tanzilal Mustaqim[3]

[1]*Faculty of Technology Management and Business, University of Tun Hussein Onn Malaysia, Malaysia*

[2]*Department of Science Engineering, Sungkyuwan University, South Korea*

[3]*Department of Computer Science, State University of Semarang, Indonesia*

Abstract: Bankruptcy analysis is needed to anticipate bankruptcy. Errors in predicting bankruptcy often cause bankruptcy. Machine learning with high accuracy to analyze reversal must continuously improve its accuracy. Many machine learning models have been applied to predict bankruptcy. However, model improvisation is still needed to improve prediction accuracy. We propose a combination model to improve the accuracy of bankruptcy prediction based on a genetic algorithm-support vector machine (GA-SVM) and stacking ensemble method. This study uses the Taiwanese Bankruptcy dataset from the Taiwan Economic Journal. Then, we implement a synthetic minority over-sampling technique for handling imbalanced datasets. We select the best feature using GA-SVM, adopt a new strategy by stacking the classifier, and use extreme gradient boosting as a meta-learner. The results show superior accuracy obtained by the stacking model-based GA-SVM with an accuracy of 99.58%. The accuracy obtained is higher than just applying a single classifier. Thus, this study shows that the proposed method can predict bankruptcy with superior accuracy.

Keywords: bankruptcy prediction, Taiwanese Bankruptcy, genetic algorithm, stacking ensemble, SMOTE

1. Introduction

Financial failure or bankruptcy is facing multiple crises that could devastate the global economy (Abdillah, 2020). Companies, small business owners, conventional marketplaces, and the government might all experience a detrimental influence. Academic scholars, government officials, investors, and business people are attempting to pinpoint the factors that lead to bankruptcy (Liang et al., 2016). The cost of raw materials, employee pay, company competitiveness, and managerial incompetence are just a few factors that can lead to bankruptcy. Every business owner is susceptible to bankruptcy-related reasons. It may influence entrepreneurs at every level (Blazy & Stef, 2020). Managing funds and lowering economic credit risk can enhance credit risk assessment (Umar et al., 2021). To safeguard their finances from bankruptcy, investors, managers, shareholders, and governments may all benefit from the information provided by bankruptcy assessments. Financial weak spots can be identified, and early warning signs can be supplied through bankruptcy research. However, a few benefits of bankruptcy study include lowering the cost of credit analysis, financial monitoring, and collection rates.

*Corresponding Author: Much Aziz Muslim, Faculty of Technology Management and Business, University of Tun Hussein Onn Malaysia, Malaysia. Email: a212muslim@yahoo.com

The financial statements contain information such as the company's cash flow and net income. Cash flow is lower than the target than the cost of capital issued, and unpaid debt can lead to liquidity, where the company can be dissolved at the discretion of the local country. Even the failure of the state economy can accelerate the company's bankruptcy if the company only relies on state subsidies (Foerster et al., 2017).

Several factors may be considered to determine the causes of bankruptcy, including the economy, changes in the price of commodities, the surplus and deficit of the nation, and even the strength of the local currency. Socially, changes in community customs will affect the company's target market. If the company can't adjust and adapt, it will lead to liquidation. The company's use of technology and media must be exceeded, and it will be lost if it doesn't improve its systems. This can affect the maintenance of the company's equipment; otherwise, having an integrated system will only make it difficult for the company. Governmental regulations are about removing subsidies and tariffs on the movement of export and import sales (Hickel et al., 2022). Ineffective business management may also result in insolvency (Kücher et al., 2020).

Financial bankruptcy analysis is needed to improve financial management (Muslim & Dasril, 2021). It requires measures and indicators to analyze financial bankruptcy (Kozlovskyi et al., 2019). Bateni and Asghari (2020) initiated a bankruptcy analysis of 158 companies and formed six groups of financial ratios (FRs) as indicators of bankruptcy prediction, namely cash flow to total debt, net income to total assets, accumulation of current and long-term liabilities to total assets, current ratio, capital employment to total assets, and the interval without credit. Almamy et al. (2016) provides further research on financial bankruptcy and found five groups of financial ratios used as indicators of bankruptcy prediction: cash flow to total debt, net income to total assets, total debt to total assets, working capital, and the current ratio. Other ratios that can support the need for analysis are liquidity and leverage (Jumanto et al., 2023). The bankruptcy prediction indicator is then referred to as the z-score. Research by Liang et al. (2016) explained that two critical factors in analyzing bankruptcy are FRs and Corporate Governance Indicators (CGIs). FRs factors, are solvency, profitability, cash flow ratio, capital structure ratio, turnover, and growth. CGI factors are board structure, ownership structure, cash flow rights, and retained experts. These two factors will develop the model's performance even though the dimensions obtained are high.

Financial bankruptcy indicators will form a pattern, which is not enough if it only produces an analysis of human assumptions (Muslim & Dasril, 2021). Statistical analysis and machine learning are examples of models to minimize bankruptcy (Lin et al., 2019). Financial distress can be discriminated against by combining FRs, and specific ratios can be determined using multiple discriminant analyses. However, it is not strong enough to prove that only using MDA can produce the best results (Liang et al., 2020). Bankruptcy analysis is like the classification model, taken from statistical data provided by the company to map characteristics and indicators of the causes of bankruptcy.

Feature selection is used according to the type of data object owned. Data that have a label is called supervised. The wrapper method is one variation of the supervised feature selection method that can reduce high dimensions (El Aboudi & Benhlima, 2016). The genetic algorithm (GA) has natural dimensions characteristics by applying the rules of selection, crossover, and mutation. GA characteristic rules can solve nonlinear optimization problems and form discriminant classifiers. GA can be used as a wrapper in the feature selection step. The wrapper method will iterate based on the desired generation formation (El Aboudi & Benhlima, 2016). The model object used is the SVM; SVM can solve complex problems such as high data dimensions and nonlinear data (Zhou et al., 2019).

The results of the subset of features and data that have reached the specified generation (stop criteria) will be trained using the ensemble method. The ensemble method will train the data in parallel and produces accuracy values (Dong et al., 2020). One of the algorithms applied to the ensemble method is stacking (Jayapermana et al., 2022). The stacking algorithm can train data based on the machine learning algorithm used. The stacking algorithm will divide the training into the base and meta-learners (Dou et al., 2020). The data will be trained using a predetermined algorithm at the base learner stage. This study uses three machine learning algorithms as a base learner, namely a decision tree (Zaman et al., 2020), a k-nearest neighbor (Wang & Liu, 2021), and a light gradient boosting machine (Ke et al., 2017). At the meta-learner stage, extreme gradient boosting is used to predict bankruptcy (Khoirunnisa et al., 2021; Muslim & Dasril, 2021).

Research by Zelenkov et al. (2017) used the ensemble classifier even though they did not use the filtering method. Barboza et al. (2017) have applied ensemble learning AdaBoost to reduce the error rate in selecting FRs attributes. Research by Liang et al. (2016) used GA as feature selection to reduce data dimensions and used SVM with linear kernel parameters to find the best features. Muslim and Dasril (2021) use the extreme gradient boosting algorithm as feature selection and ensemble stacking with the base learner, namely k-nearest neighbor, decision tree, gradient boosting trees, and random forest. The meta-learner used is a light gradient boosting machine. Many studies have been conducted using bankruptcy databases, including Li et al. (2017). It compiles information on Italy's small- and medium-sized firms and uses McNemar validation to examine insolvency. The bankruptcy dataset for Poland is utilized by Muslim and Dasril (2021) and Fernández et al. (2018), the Taiwanese dataset is regularly used by Liang et al. (2016), and Lin et al. (2019) examine the comparison between the single and ensemble learning-based bankruptcy.

Research related to bankruptcy prediction also implements the ensemble technique that has been carried out by Muslim and Dasril (2021) using Polish Bankruptcy data, Pisula (2020) using Poland Bankruptcy data, and Guo et al. (2022) using the Taiwanese Bankruptcy dataset. Each of these studies proposes the ensemble learning method with different meta-learner variations and feature selection methods. Therefore, this research is proposed based on similar research references.

Previous research on bankruptcy prediction only focused on increasing the classification model's accuracy, ignored the excessive number of variable data dimensions, and did not make efforts to balance class data. Although the classification model applied is taken from several references to bankruptcy prediction research that has been done, namely the stacking ensemble method, this research adds a method for handling unbalanced data and a wrapper technique for selecting the best features. Improving this method is done as an effort to improve the accuracy of bankruptcy prediction performance.

In this study, we improve the bankruptcy prediction accuracy based on a genetic algorithm-support vector machine (GA-SVM) and stacking ensemble method on the Taiwanese Bankruptcy dataset from the Taiwan Economic Journal. The rest of the manuscript is organized as follows. Section 2 reviews the related works to bankruptcy, feature selection model, and classification model. Section 3 presents the GA-SVM method, stacking ensemble model, and other machine learning models. Section 4 analyses and discusses the results. Section 5 conclusion.

2. Literature Review

Studies on bankruptcy prediction by applying a classification model are based on statistical data provided by financial companies, such as those conducted by du Jardin (2016), He et al. (2018), and Kliestik et al. (2020). Classification algorithms can solve classification problems, such as multivariant discriminant analysis, logistic regression, neural networks, support vector machine (SVM), and ensemble method (Qu et al., 2019). The use of other algorithms that use the characteristics of the neural network, such as the recurrent neural network (Chong et al., 2017) and convolutional neural network (Hosaka, 2019), have been developed to analyze financial and management topics (Qu et al., 2019). Using feature selection, high indicator dimensions may be decreased to increase prediction accuracy (Bouktif et al., 2018; Pampouktsi et al., 2023). Linero (2018) mentions that high dimensions can degrade the performance of the resulting accuracy prediction, such as in the decision tree and Naïve Bayes because the data's attributes are irrelevant to the algorithm. Feature selection can reduce the high dimensions obtained from the combination of FRs and CGIs indicators (Liang et al., 2016).

The feature selection method has been popularly optimized for classification problems (Dessì & Pes, 2015; Tran et al., 2018; Zhang et al., 2018). Research by Liang et al. (2016) examined the effect of the filter and wrapper-based feature selection methods in bankruptcy prediction. They found no optimal combination of feature selection techniques and classification methods in experimental datasets. The empirical results indicate that the proposed wrapper method performs better than traditional feature selection models regarding prediction accuracy.

A dataset with large dimensions, like the Taiwanese dataset compiled by Liang et al. (2016), is caused by using variables to determine financial insolvency that might total more than 20. Large dataset size can make it difficult to compile bankruptcy prediction analyses; hence, feature selection techniques such as GA (Liang et al., 2016; Lin et al., 2019; Prasetiyo et al., 2021), extreme gradient boosting (Muslim & Dasril, 2021), and SVM (Lin et al., 2019) are frequently utilized.

The use of feature selection has aided efforts to increase prediction accuracy, yet it is insufficient. Single classifier GA (Ajani et al., 2021; Li et al., 2017; Lin et al., 2019), logistic regression (Li et al., 2017), k-nearest (Noori et al., 2017), SVM (Liang et al., 2016; Lin et al., 2019; Prasetiyo et al., 2021; Qu et al., 2019), and light gradient boosting machine (Zelenkov et al., 2017) were used for dataset training utilizing machine learning. Any classification method can be stacked using a stacking ensemble to improve accuracy (Muslim & Dasril, 2021).

Research by Brenes et al. (2022) improved a multilayer perceptron model to predict bankruptcy with a dataset of Taiwanese firms to obtain an accuracy of 86.06%. A similar study combined the stacking ensemble and XGBoost feature selection algorithms to obtain an accuracy of 97% (Muslim & Dasril, 2021). However, from previous research, it is still necessary to improvise methods to obtain maximum accuracy. So, based on the related work description above, the researcher found a strategy on how to find a method to predict bankruptcy with high accuracy from previous studies.

3. Research Methodology

3.1. Research design

The research design was made to explain the research flow comprehensively. The workflow generally describes three stages: the preprocessing stage, the feature selection stage, and the data training stage. The study ended with a validation test using a confusion matrix and k-fold cross-validation. The visualization of the proposed method is shown in Figure 1.

3.1.1. Data description

A random sampling method was employed for the study. The sample for the study consisted of 22 first-year in-service postgraduate science teachers from one of the Colleges of Education in Bhutan. Thirteen male (59%) teachers and nine female (40.9%) teachers participated in the study. The sample comprised 7 biology teachers (31.8%), 10 chemistry teachers (45.5%), and 5 physics teachers (22.7%). Since the participation in this study was purely voluntary, only 22, out of 39 in-service postgraduate science teachers participated. The overall response rate was recorded at 56%.

3.1.2. Preprocessing stage

The preprocessing stage is the stage that aims to process the dataset before being trained using the algorithm (Li et al., 2017). The dataset detected abnormality, noise, and balance of the dataset. An unbalanced dataset and low accuracy can affect the training process (Fernández et al., 2018). The preprocessing stage used in this study is the synthetic minority over-sampling technique. The synthetic minority over-sampling technique can be used for noisy datasets so that the synthetic minority over-sampling technique will form balanced data based on the distribution of the nearest k value (Prasetiyo et al., 2021).

3.1.3. Feature selection stage

The feature selection stage aims to reduce the dimensions of the dataset based on the features used. Feature selection combines GA and SVM. A wrapper combines the two algorithms (El Aboudi & Benhlima, 2016). The GA parameters use the stopping criteria based on the formed generation iteration. The generation used is 20, the crossover rate is 0.7, the population size is 70, and the mutation rate is 0.01 (Liang et al., 2016). An illustration of the GA is shown in Figure 2.

SVM is used as an evaluator for each iteration. SVM evaluator applies to assess the level of accuracy of features after selection. The SVM model in pseudocode is as follows (Ajani et al., 2021):

Algorithm 1. SVM Model

Require: X and y loaded with labeled data, α 0
1: C <= some value (use 100, for example)
2: Kernel ⇐ formula (RBF, for example)
3: **repeat**
4: **for all** $\{x_i, y_i\},\{x_j, y_j\}$ **do**
5: Optimize α_i and α_j
6: **end for**
7: **until** no changes in α or other resource constraint criteria are met
Ensure: Retain only the support vector $(\alpha_i > 0)$

3.1.4. Stacking ensemble learning

One of the algorithms applied to the ensemble method is stacking (Jayapermana et al., 2022). The stacking algorithm can train data based on the machine learning algorithm used. The ensemble technique combines multiple models to produce a more accurate model. The ensemble method can significantly lower misclassification and raise the efficiency of a single classification model (Abdar et al., 2020). To produce a more accurate model,

Figure 1
The proposed method of bankruptcy prediction with GA-SVM and stacking

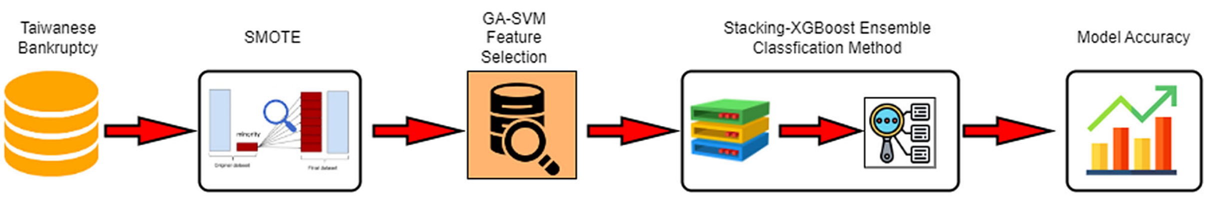
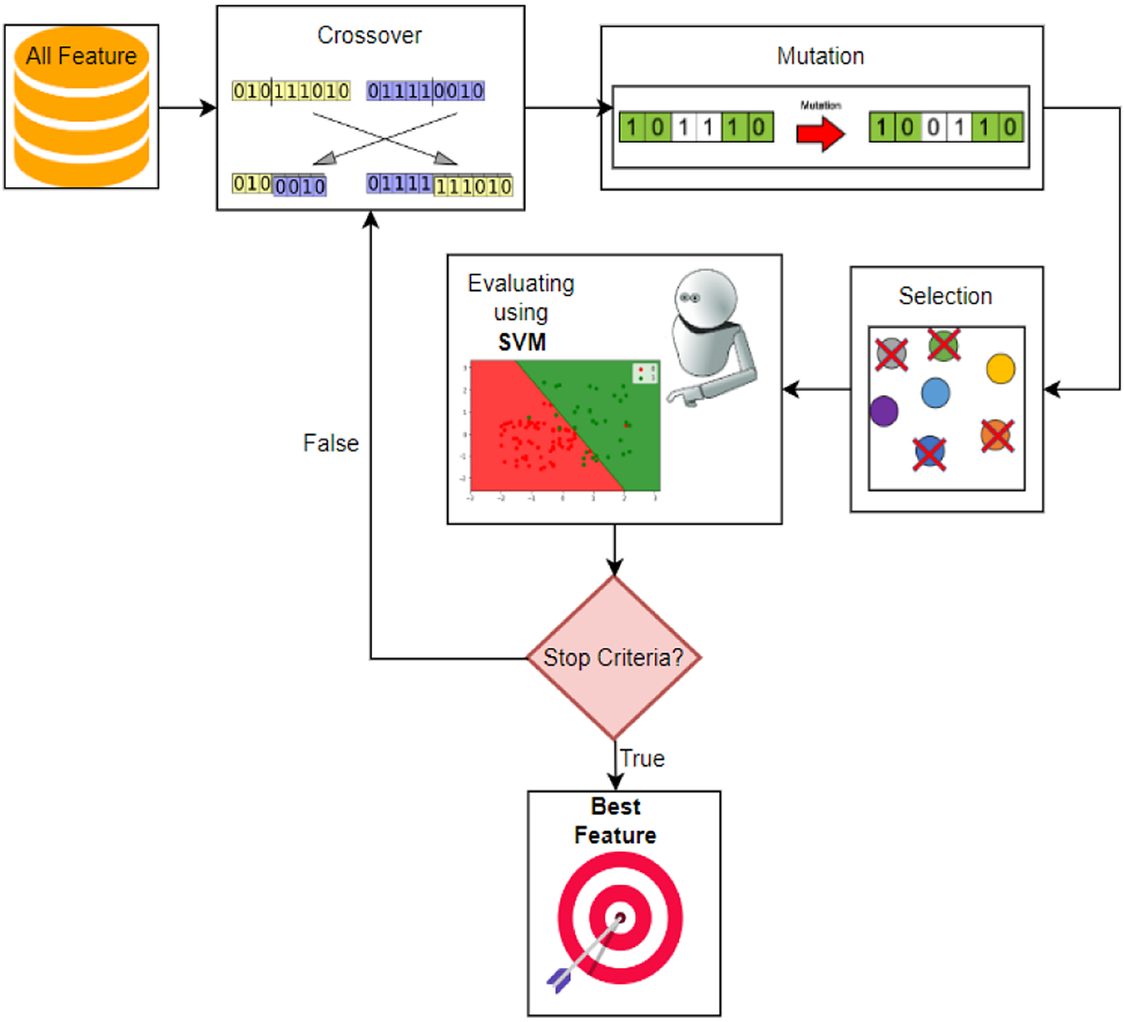


Figure 2
Illustration of the feature selection process of GA-SVM



the ensemble technique integrates various sets of models. The ensemble technique can greatly reduce misclassification and boost a single classifier's effectiveness. The base model and meta-learner are the two main concepts that the algorithm suggests.

The fundamental concept is to use the output of the meta-learner as the final prediction result after merging the strategies and to train the meta-learner by using the prediction result of the base model as a new feature matrix. The multilayer stacking ensemble learning model uses the output from one layer as the input for the one below it. However, as the number of layers rises, the model becomes more complex, and the training rate declines. This study uses three machine learning algorithms as a base learner, namely a decision tree (Zaman et al., 2020), a k-nearest neighbor (Wang & Liu, 2021), and a light gradient boosting machine (Ke et al., 2017). At the meta-learner stage, extreme gradient boosting is used to predict bankruptcy (Khoirunnisa et al., 2021; Muslim & Dasril, 2021).

4. Result and Discussion

The dataset used in the study shows that the comparison of the target classification label "Bankrupt?" between label 1 and label 0 has a ratio of 96.77%:3.23%. Based on the original dataset, the target label is 0 for as many as 6599 instances, and the target label 1 is 220 instances. The target label comparison visualization is in Figure 3.

The imbalanced data target in Figure 3 needs to be improved because it will affect the accuracy of the data training. Repairing imbalanced data using resamples data using the nearest k value in the distribution of the dataset. The target label "Bankrupt?" 0 is called the majority label, while the target label "Bankrupt?" 1 is

Figure 3
Original dataset label target of Taiwanese Bankruptcy dataset

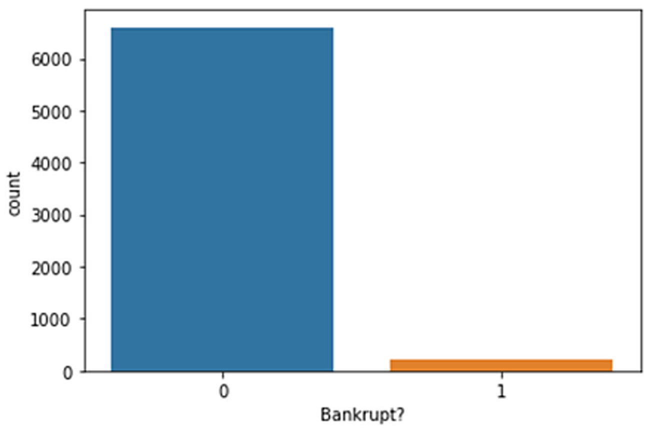
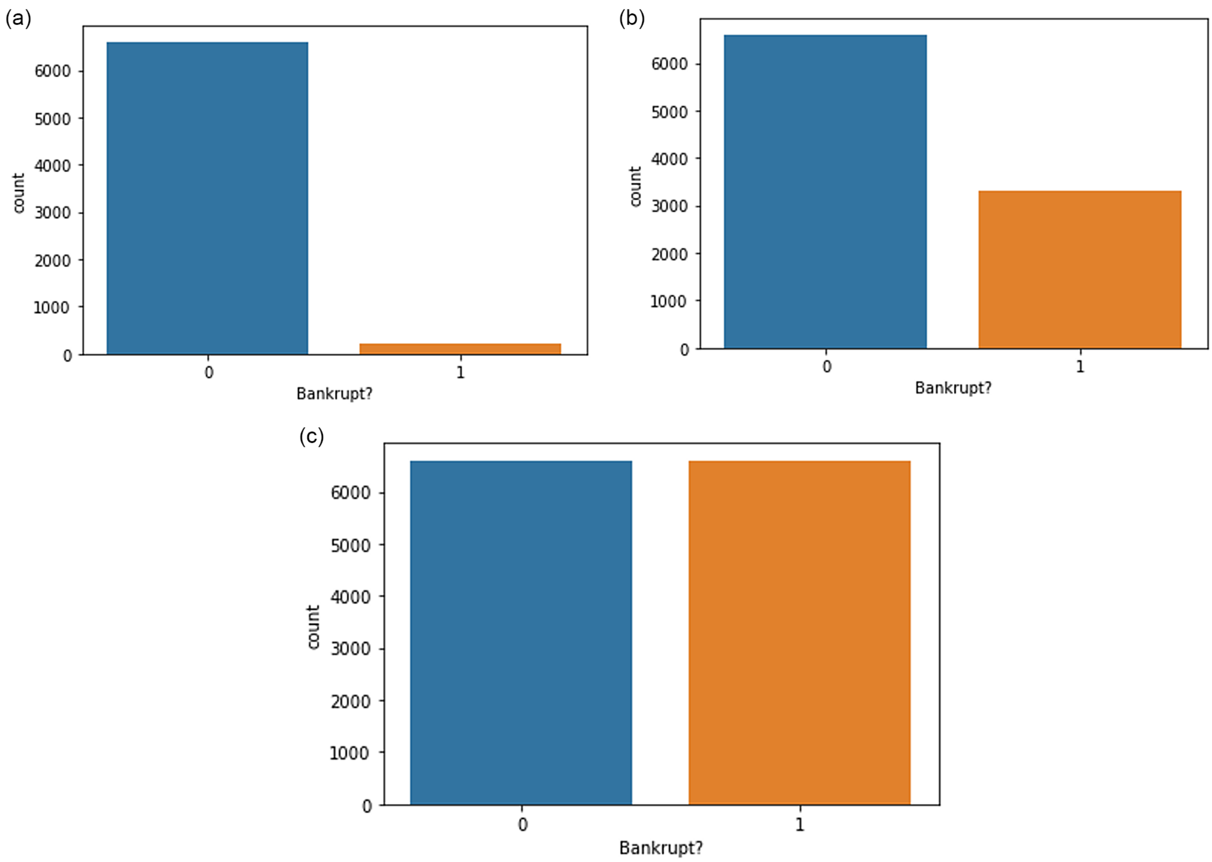


Figure 4
The use of the synthetic minority oversampling technique method. Figure (a) is the original data, (b) resampling 50% of the majority label, and (c) resampling 100% of the majority label



called the minority label. A minority label will be applied to resampling the dataset using a synthetic over-sampling technique (Prasetiyo et al., 2021). The resampling process that is applied starts from 50% of the dataset from the majority label, then the resampling of the dataset used is 100% from the majority label. Visualization of the synthetic minority over-sampling technique carried out is in Figure 4.

The number of resampling results formed based on the specified percentage is in Table 1.

The results of the new dataset used for data training are minority labels resampled by 100% of the majority labels. Thus, the dataset used to conduct data training is balanced data. Balanced data is processed using GA-SVM to reduce used features. In Noori et al.'s (2017) research, three features of accuracy were obtained at 91.00% using fNIRS signal data, while Gokulnath and Shantharajah's (2019) research obtained 7 features from heart disease data and resulted in an accuracy of 83.70%. In this research, the GA-SVM approach is used as feature selection with Taiwanese Bankruptcy data obtained 43 instances with an accuracy of 92.05%, as shown in Table 2.

The best features selected by GA-SVM are 43 features out of 92 native features. The result of the training data process produces different levels of accuracy. The algorithms are a decision tree, k-nearest neighbor, and light gradient boosting machine. In addition to using a single classifier, the data training model uses stacking. The result of the comparison of training data testing using a single classifier that has been determined and the stacking algorithm using meta-learner extreme gradient boosting is in Table 3 (Hou et al., 2021).

The result of data training without feature selection GA-SVM using a decision tree is 95.15%, using a k-nearest neighbor produces 90.8%, and light gradient boosting machine produces 98.37%. The result of data training using GA-SVM feature selection using a decision tree is 96.02%, the k-nearest neighbor produces 91.25%, and the light gradient boosting machine produces 98.45%. To the results of research from Tao et al. (2018).

Table 1
Comparison of the original target dataset and normalized target dataset

Label "Bankrupt?"	Original data	Resampling percentage from majority label 50%	Resampling percentage from majority label 100%
0	6599	6599	6599
1	220	3299	6599

Table 2
Comparison feature after being selected using GA-SVM

All features	Best features
95	43

Table 3
Comparison of single classifier without GA-SVM and using GA-SVM

	Accuracy (%)	
Algorithm	Without GA-SVM	Using GA-SVM
Decision tree	95.15	96.02
K-nearest neighbor	90.80	91.25
Light gradient boosting machine	98.37	98.45

Table 4
Comparison of single classifier and ensemble classifier

Algorithm	Category	Accuracy (%)
Decision tree	Single classifier	96.02
K-nearest neighbor	Single classifier	91.25
Light gradient boosting machine	Single classifier	98.45
Stacking-extreme gradient boosting	Ensemble classifier	99.58

Figure 5
Comparison of single classifier without GA-SVM and with GA-SVM

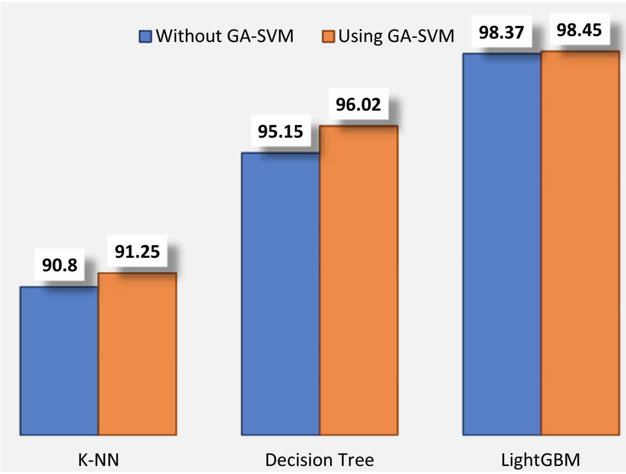
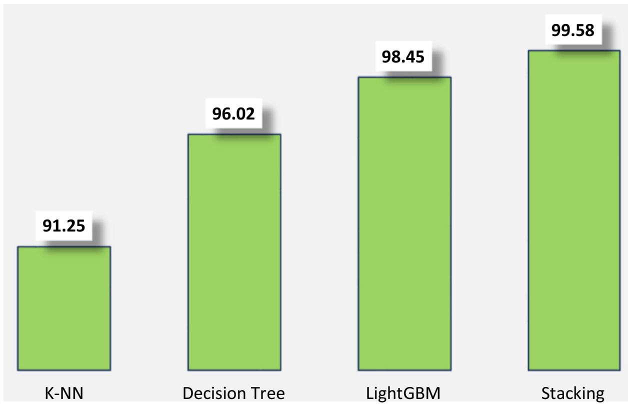


Figure 6
Comparison of single classifier and stacking



The results obtained from feature selection using GA-SVM on each single classifier have increased accuracy. Can be seen as in Figure 5.

The ensemble stacking method uses the decision tree, k-nearest neighbor, and light gradient boosting machine as a base learner. A novel dataset that uses meta-learning to categorize was obtained from the base learner. The meta-learner used in the ensemble stacking method is extreme gradient boosting classification. The results are shown in the comparison between single classifier (decision tree, k-nearest neighbor, and light gradient boosting machine) and ensemble classifier (extreme gradient boosting with stacked decision tree, k-nearest neighbor, and light gradient boosting machine) Table 4.

An accuracy of 99.58% is achieved by stacking a decision tree, k-nearest neighbor, and light gradient boosting machine with an ensemble classifier as a base learner and an extreme gradient boosting as a meta-learner. The comparison visualization of comparison single classifier and stacking is in Figure 6.

In research conducted by Muslim and Dasril (2021) using Polish Bankruptcy data and ensemble stacking, the results obtained were 97%, while research conducted by Pisula (2020) using Poland Bankruptcy data ensemble stacking resulted in an accuracy of 98.1%. In addition, a similar experiment by Guo et al. (2022) using the Taiwanese Bankruptcy dataset and testing the bagging ensemble model obtained an accuracy of 86.63%. Thus, it can be shown that the proposed stacking ensemble learning method with Taiwanese Bankruptcy data is superior.

5. Conclusion

Analysis of bankruptcy prediction using machine learning is done through research. The research used the GA-SVM machine learning algorithm as feature selection, a single classifier as a base learner, and extreme gradient boosting as a meta-learner. The single classifier is used as a decision tree, k-nearest neighbor, and light gradient boosting machine. The results of the dataset training after the attributes are selected using GA-SVM through a single classifier: the decision tree produces 96.02%, the k-nearest neighbor produces 91.25%, and the light gradient boosting machine produces 98.45%. Dataset training using the ensemble stacking method resulted in 99.58%. It is proven that using a stacking ensemble with meta-learner extreme gradient boosting can increase the accuracy value in bankruptcy classification.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Much Aziz Muslim is an editorial board member for *Journal of Data Science and Intelligent Systems* and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

References

Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters*, *132*, 123–131. https://doi.org/10.1016/j.patrec.2018.11.004.

Abdillah, L. A. (2020). FinTech E-commerce payment application user experience analysis during COVID-19 pandemic. *arXiv Preprint: 2012.07750*. https://doi.org/10.48550/arXiv.2012.07750

Ajani, T. S., Imoize, A. L., & Atayero, A. A. (2021). An overview of machine learning within embedded and mobile devices – Optimizations and applications. *Sensors*, *21*(13), 4412. https://doi.org/10.3390/s21134412

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance*, *36*, 278–285. https://doi.org/10.1016/j.jcorpfin.2015.12.009

Barboza, F., Kimura, H., & 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

Bateni, L., & Asghari, F. (2020). Bankruptcy prediction using logit and genetic algorithm models: A comparative analysis. *Computational Economics*, *55*, 335–348. https://doi.org/10.1007/s10614-016-9590-3

Blazy, R., & Stef, N. (2020). Bankruptcy procedures in the post-transition economies. *European Journal of Law and Economics*, *50*, 7–64. https://doi.org/10.1007/s10657-019-09634-5

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636. https://doi.org/10.3390/en11071636

Brenes, R. F., Johannssen, A., & Chukhrova, N. (2022). An intelligent bankruptcy prediction model using a multi-layer perceptron. *Intelligent Systems with Applications*, *16*, 200136. https://doi.org/10.1016/j.iswa.2022.200136

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Dessì, N., & Pes, B. (2015). Similarity of feature selection methods: An empirical study across data intensive classification tasks. *Expert Systems with Applications*, *42*(10), 4632–4642. https://doi.org/10.1016/j.eswa.2015.01.069

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241–258. https://doi.org/10.1007/s11704-019-8208-z

Dou, J., Yunus, A. P., Bui, D. T., Merghadi, A., Sahana, M., Zhu, Z., . . . , & Pham, B. T. (2020). Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides*, *17*, 641–658. https://doi.org/10.1007/s10346-019-01286-5

du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, *254*(1), 236–252. https://doi.org/10.1016/j.ejor.2016.03.008

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. In *2016 International Conference on Engineering & MIS*, 1–5. https://doi.org/10.1109/ICEMIS.2016.7745366

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/10.1613/jair.1.11192

Foerster, S., Tsagarelis, J., & Wang, G. (2017). Are cash flows better stock return predictors than profits? *Financial Analysts Journal*, *73*(1), 73–99. https://doi.org/10.2469/faj.v73.n1.2

Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, *22*, 14777–14787. https://doi.org/10.1007/s10586-018-2416-4

Guo, Y., Mei, J., Pan, Z., Liu, H., & Li, W. (2022). Adaptively promoting diversity in a novel ensemble method for imbalanced credit-risk evaluation. *Mathematics*, *10*(11), 1790. https://doi.org/10.3390/math10111790

He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, *98*, 105–117. https://doi.org/10.1016/j.eswa.2018.01.012

Hickel, J., Dorninger, C., Wieland, H., & Suwandi, I. (2022). Imperialist appropriation in the world economy: Drain from the global south through unequal exchange, 1990–2015. *Global Environmental Change*, *73*, 102467. https://doi.org/10.1016/j.gloenvcha.2022.102467

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287–299. https://doi.org/10.1016/j.eswa.2018.09.039

Hou, H., Chen, X., Li, M., Zhu, L., Huang, Y., & Yu, J. (2021). Prediction of user outage under typhoon disaster based on multi-algorithm Stacking integration. *International Journal of Electrical Power & Energy Systems*, *131*, 107123. https://doi.org/10.1016/j.ijepes.2021.107123

Jayapermana, R., Aradea, A., & Kurniati, N. I. (2022). Implementation of stacking ensemble classifier for multi-class classification of COVID-19 vaccines topics on Twitter. *Scientific Journal of Informatics*, *9*(1), 8–15.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . , & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *31st Conference on Neural Information Processing Systems*.

Khoirunnisa, O., Pertiwi, D. A. A., Dianti, E. N., Fattah, A. M. M., & Budiman, K. (2021). Improvement business process model and notation on the drink distribution industries using six core element. *Journal of Soft Computing Exploration*, *2*(2), 99–106. https://doi.org/10.52465/joscex.v2i2.46

Kliestik, T., Valaskova, K., Lazaroiu, G., Kovacova, M., & Vrbka, J. (2020). Remaining financially healthy and competitive: The role of financial predictors. *Journal of Competitiveness*, *12*(1), 74–92. https://doi.org/10.7441/joc.2020.01.05

Kozlovskyi, S., Poliakov, B., Lavrov, R., & Ivanyuta, N. (2019). Management and comprehensive assessment of the probability of bankruptcy of Ukrainian enterprises based on the methods of fuzzy sets theory. *Problems and Perspectives in Management*, *17*(3), 370–381. http://dx.doi.org/10.21511/ppm.17(3).2019.30

Kücher, A., Mayr, S., Mitter, C., Duller, C., & Feldbauer-Durstmüller, B. (2020). Firm age dynamics and causes of corporate bankruptcy: Age dependent explanations for business failure. *Review of Managerial Science*, *14*, 633–661. https://doi.org/10.1007/s11846-018-0303-2

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys*, *50*(6), 94. https://doi.org/10.1145/3136625

Liang, D., Lu, C. C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, *252*(2), 561–572. https://doi.org/10.1016/j.ejor.2016.01.012

Liang, D., Tsai, C. F., Lu, H. Y., & Chang, L. S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research*, *120*, 137–146. https://doi.org/10.1016/j.jbusres.2020.07.052

Lin, W. C., Lu, Y. H., & Tsai, C. F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems*, *36*(1), e12335. https://doi.org/10.1111/exsy.12335

Linero, A. R. (2018). Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, *113*(522), 626–636. https://doi.org/10.1080/01621459.2016.1264957

Muslim, M. A., & Dasril, Y. (2021). Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning. *International Journal of Electrical and Computer Engineering*, *11*(6), 5549–5557. https://doi.org/10.11591/ijece.v11i6.pp5549-5557

Jumanto, J., Muslim, M. A., Dasril, Y., & Mustaqim, T. (2023). Accuracy of Malaysia public response to economic factors during the Covid-19 pandemic using vader and random. *Journal of Information System Exploration and Research*, *1*(1), 49–70. https://doi.org/10.52465/joiser.v1i1.104

Noori, F. M., Naseer, N., Qureshi, N. K., Nazeer, H., & Khan, R. A. (2017). Optimal feature selection from fNIRS signals using genetic algorithms for BCI. *Neuroscience Letters*, *647*, 61–66. https://doi.org/10.1016/j.neulet.2017.03.013

Pampouktsi, P., Avdimiotis, S., Maragoudakis, M., Avlonitis, M., Samantha, N., Hoogar, P., …, & Rono, W. (2023). Techniques of applied machine learning being utilized for the purpose of selecting and placing human resources within the public sector. *Journal of Information System Exploration and Research*, *1*(1), 1–16. https://doi.org/10.52465/joiser.v1i1.91

Pisula, T. (2020). An ensemble classifier-based scoring model for predicting bankruptcy of polish companies in the Podkarpackie Voivodeship. *Journal of Risk and Financial Management*, *13*(2), 37. https://doi.org/10.3390/jrfm13020037

Prasetiyo, B., Muslim, M. A., & Baroroh, N. (2021). Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique. *Journal of Physics: Conference Series*, *1918*, 42002. https://doi.org/10.1088/1742-6596/1918/4/042002

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *162*, 895–899. https://doi.org/10.1016/j.procs.2019.12.065

Tao, P., Sun, Z., & Sun, Z. (2018). An improved intrusion detection algorithm based on GA and SVM. *IEEE Access*, *6*, 13624–13631. https://doi.org/10.1109/ACCESS.2018.2810198

Zhou, T., Lu, H., Wang, W., & Yong, Y. (2019). GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Applied Soft Computing*, *75*, 323–332. https://doi.org/10.1016/j.asoc.2018.11.001

Tran, C. T., Zhang, M., Andreae, P., Xue, B., & Bui, L. T. (2018). Improving performance of classification on incomplete data using feature selection and clustering. *Applied Soft Computing*, *73*, 848–861. https://doi.org/10.1016/j.asoc.2018.09.026

Umar, M., Ji, X., Mirza, N., & Naqvi, B. (2021). Carbon neutrality, bank lending, and credit risk: Evidence from the Eurozone. *Journal of Environmental Management*, *296*, 113156. https://doi.org/10.1016/j.jenvman.2021.113156

Wang, H., & Liu, X. (2021). Undersampling bankruptcy prediction: Taiwan bankruptcy data. *PLOS ONE*, *16*(7), e0254030. https://doi.org/10.1371/journal.pone.0254030

Zaman, M., Kaul, S., & Ahmed, M. (2020). Analytical comparison between the information gain and Gini index using historical geographical data. *International Journal of Advanced Computer Science and Applications*, *11*.

Zelenkov, Y., Fedorova, E., & Chekrizov, D. (2017). Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications*, *88*, 393–401. https://doi.org/10.1016/j.eswa.2017.07.025

Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, *106*, 64–85. https://doi.org/10.1016/j.dss.2017.12.001

How to Cite: Aziz Muslim, M., Dasril, Y., Javed, H., Alamsyah, Jumanto, Faqih Abror, W., Agustina Pertiwi, D. A., & Mustaqim, T. (2024). An Ensemble Stacking Algorithm to Improve Model Accuracy in Bankruptcy Prediction. *Journal of Data Science and Intelligent Systems*, *2*(2), 79–86. https://doi.org/10.47852/bonviewJDSIS3202655