Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning

Much Aziz Muslim[1], Yosza Dasril[2]
[1]Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Malaysia
[1]Department of Computer Science, Universitas Negeri Semarang, Indonesia
[2]Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia

| Article Info	ABSTRACT
	Company bankruptcy is often a very big problem for companies. The impact of bankruptcy can cause losses to elements of the company such as owners, investors, employees, and consumers. One way to prevent bankruptcy is to predict the possibility of bankruptcy based on the company's financial data. Therefore, this study aims to find the best predictive model or method to predict company bankruptcy using the dataset from Polish companies bankruptcy. The prediction analysis process uses the best feature selection and ensemble learning. The best feature selection is selected using feature importance to XGBoost with a weight value filter of 10. The ensemble learning method used is stacking. Stacking is composed of the base model and meta learner. The base model consists of K-nearest neighbor, decision tree, support vector machines (SVM), and random forest, while the meta learner used is LightGBM. The stacking model accuracy results can outperform the base model accuracy with an accuracy rate of 97%.
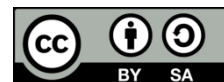
Corresponding Author:

Much Aziz Muslim
Faculty of Technology Management
Universiti Tun Hussein Onn Malaysia
Batu Pahat, Johor, Malaysia
Email: a212muslim@mail.unnes.ac.id

1. INTRODUCTION

Predicting company bankruptcy is one of the most important parts of management science problems. The main purpose of this prediction is to categorize companies that are safe and unsafe or bankrupt [1]. In addition, the wrong decision-making in financial institutions that are in financial difficulty or distress is experienced by many social costs such as owners or shareholders, managers, government and others. Therefore, the prediction of company bankruptcy has become a special concern among industrial practitioners as well as academics or researchers [2]-[5].

Nowadays, machine learning techniques [6] and artificial intelligence [7] computation have been widely used by researchers to solve bankruptcy prediction problems such as support vector machines (SVM) [8]-[16], decision trees [17]-[23], artificial neural networks (ANN) [24]-[31] and discussion with systematic literature review technique [32]-[37]. Meanwhile, improvement in machine learning techniques through various strategies has also been carried out such as boosting improvement based on feature selection known as FS-Boosting is proven to have good performance as a learner and has higher accuracy and diversity based on two selected company bankruptcy data sets [38]. The combination of SVM and ANN integrated with dropout, auto-encoder proved to produce better accuracy than logistic regression, genetic algorithm and

inductive learning [39]. A hybrid approach based on synthetic minority over-sampling technique known as the SMOTE technique with the ensemble learning method, i.e. Boosting, Bagging, Naive Bayes, ANN, Random forest, Rotation forest and diverse ensemble creation by oppositional relabeling of meaningful training examples (DECORATE) are proven to efficiently improve performance parameters such as accuracy, AUC, error types 1 and 2, G-mean through the collected data set of Spanish companies [40]. The integration approach of SVM proportions, boosting and bagging in an ensemble strategy called Bagged-pSVM and Boosted-pSVM which is based on a learning perspective with label proportions where unlabeled learning data are provided with different bags and only given a bag based on the proportion of instances level with particular classes. This approach is proposed to overcome a large number of instance-level labeled learning data [41]. The hybrid of SMOTE-edited nearest neighbor (SMOTE-ENN) as over-sampling technique and CBoost algorithm as cost-sensitive learning or predictive model. This hybrid produces the best performance of existing learning techniques [42]. Reducing the unbalanced class of bankruptcy data sets using over-sampling or SMOTE techniques then ANN as a predictive model. This concept resulted in significant performance than the ANN and weak learners trained in the AUC section [43]. Borderline synthetic minority over-sampling technique (BSM) and stacked auto-encoder (SAE) based on the Soft-max classifier are proposed to solve the unbalanced classification of company bankruptcy prediction problems. This combination approach is considered more efficient than the combination of BSM with machine learning techniques and machine learning techniques without over-sampling [44].

At the same time, the process of running the company's business produces financial data that can be used to predict bankruptcy [45]. The latest discussion regarding bankruptcy prediction focuses on feature selection [33]. Company financial data such as sales, profit and asset data affect the analysis process of bankruptcy predictions. The resulting company financial data has many features so that the best feature analysis process is needed to improve the quality of predictions. Two types feature selection based on filter and wrapper with two types classification techniques based on bagging and boosting ensemble classifier to model predictive [46]. Son *et al.* [47] used Skewness reduction for data normalization and XBoost algorithm to select features important to serve as attributes of bankruptcy predictions. The result of Son *et al.'s* method can improve predictions with an accuracy of 17% of the AUC level. Nobre [48] used the XGBoost algorithm to feature selection combined with principal component analysis (PCA) and discrete wavelet transform (DWT) to analyze bankruptcy predictions. The results of the analysis show that the method used has a return value of 49.26%

Based on previous research, increasing accuracy is the main focus in predictive studies of corporate bankruptcy. Combined approaches or improved methods are still very much needed to achieve better accuracy. Therefore, This study uses a feature analysis approach to select the best features, and combines several machine learning algorithms (stacking ensemble) to improve accuracy. XGBoost feature importance is used to select highly influential features based on the weight value of each feature during the prediction analysis process [49]. In addition to selecting the best features, this study also combines machine learning methods consisting of K-nearest neighbor, decision tree, SVM and random forest in this case called ensemble learning with the stacking method [33]. The purpose of this study was to find the highest accuracy by selecting the best selection feature and combining several machine learning methods using a stacking ensemble.

2.   THEORETICAL BACKGROUND
2.1. Boosting tree method

Boosting is a superior method in combining several basic classifications to produce an algorithm that is superior in achieving accuracy than other classification algorithms. Boosting is an additive ensemble method that works by adding new models to reduce errors made by older or existing models. Sequentially, the models are added in such a way that no possible improvement occurs. Boosted models can produce good accuracy even though the basic classification has only slightly better accuracy than random classification, so that the basic classification is considered a weak learner [50].

In a supervised learning setting, Let data-set $D = \{(x_i, y_i): x_i \in \mathbb{R}^n, y_i \in \mathbb{R}\}$ arranged of n data with m features and n labels, a boosting tree model   uses K additive functions $g_j(x)$ to predict the out put. $\hat{y}_i = G(x_i) = \sum_{j=i}^{K} g_j(x_i)$ where $g(x) = W_{s(x)}$. Clearly, $s: \mathbb{R}^m \longrightarrow T$ indicates the structure of each tree that maps a sample to the corresponding index of leaf and $W \in T$ ia a weight of leaf with T leaves. In order to learn the function set, we minimize the function of loss $L(G) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{K} R(g_j)$ where $R(g_j) = \alpha T + \beta \parallel W \parallel^2$ is a term of regularization that penalizes model complexity. The function of loss L(g) contains K- function as parameters so it is so hard to optimize directly. Instead, we optimize the additively model. Given $\hat{y}_i^t$ be $i^t h$ sample prediction at $t^t h$ iteration. We will add g_t to minimize.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{t-1} + g_t(x_i)) + R(g_t)$$

Which means that we greedily add the g_t that most improve our model for each iteration. We use approximation of second-order that uses a gradient on this intermediate function of loss $L^{(t)}$. This is the reason we name it gradient boosting algorithm as shown in algorithm 1 in Figure 1. The Xgboost [49] is an open-source library of software that gives framework of gradient boosting for C++, Java, Python, math-lab and R. It uses a gradient-boosting algorithm that results in a prediction model in the form of an ensemble of weak prediction models, which are decision trees, typically.

Algorithm 1 Gradient Boosting

1: **procedure** INPUTS
2:     $(x, y)_{i=1}^{n}$: input data
3:     K: Number of iteration
4:     $l(y, \hat{y})$: choice of loss function
5:     $g(x, \theta)$: choice of base learner
6: **procedure** ALGORITHM
7:     initialize $\widehat{G}_0 = \arg\min_\rho \Sigma_{i=1}^{N} l(y_i, \rho)$
8:     **for** $t = 1 : K$ **do**
9:         compute $\nabla G_t(x)$
10:        fit a new base learner function $g(x, \theta_t)$
11:        Determine the best gradient descent step size ρ_t; $\rho_t = \arg\min_\rho \Sigma_{i=1}^{N} l(y_i, \widehat{g}_{t-1}(x_i) + \rho g(x_i, \theta_i))$
12:        update the estimate of function : $\widehat{g}_t = \widehat{g}_{t-1} + \rho_t g(x_i, \theta_i)$

Figure 1. Gradient boosting algorithm

2.2. Stacking ensemble modeling

The stacking ensemble introduced by Wolpert [51] then formalized by Breimen [52] and theoretically validated by Van der Laan *et al.* [53] is one of the learning algorithms known as a superior learning framework based on generalizing losses. Due to its superior performance compared to other learning algorithms, Stacking ensemble has many applications for predicting company bankruptcy. As described in algorithm 2 in Figure 2. Therefore, to improve the prediction accuracy, the stacking ensemble is proposed in this study to be combined with the XGboost algorithm.

Algorithm 2 Stacking Ensemble Modeling

1: **procedure** INPUTS
2:     training data $D = (x, y)_{i=1}^{n}$
3: **procedure** ALGORITHM
   Step 1 Develop base-level models H on D
           perform k-fold cross-validation on base level model
           $h_1 = knn()$ ▷ knn:K–Nearest Neighbor
           $h_2 = dt()$ ▷ dt: Decision Tree
           $h_3 = gb()$ ▷ gb: Gradient Boosting Tree
           $h_4 = rf()$ ▷ rf: Random Forest
   Step 2 Build new data: level-one data D_1
           $D_1 = (x, y)_{i=1}^{n}$ where $x_i = \{h_1(x_i), h_2(x_i), h_3(x_i), h_4(x_i)\}$
   Step 3 Train meta-level model Φ on D_1
           perform k-fold cross-validation on meta-level model
           $\Phi = glm()$ ▷ glm: generalized linear model
           return stacking ensemble comprising of H and Φ models

---

Figure 2. Stacking ensemble modeling algorithm

3. RESEARCH METHOD

The research method of bankruptcy prediction analysis uses several stages, i.e. data collection, pre-processing data, feature importance and modeling. Generally, the research framework can be shown in Figure 3. The data-set in this study was taken publicly from Kaggle. The data-set is historical data on bankruptcy from Polish companies and has a range of years that are listed in the data-set starting from 2000 to 2012 [54]. The data-set is composed of 65 features related to the company's business continuity process. The total data rows in the data-set are 42,627 rows. The target data-set feature is in the "class" column with detailed contents, namely 0 and 1. Variable data 0 means that it is not bankrupt and vice versa in data variable 1 indicates bankruptcy. Data pre-processing means normalizing data sets that do not support the analysis process [47], [55]. Data that do not support the analysis process are repetitive data, blank data and abnormal data. Features that are not related to the analysis process in the data set will be normalized [56], [57]. The data-preprocessing method in this study is data scaling. Data scaling is a method of simplifying the range of numeric data values in a data-set that has the same value [58]. Data scaling creates a balanced range of numeric data. Importance features are selected based on the calculation of the XGBoost algorithm [48]. The method of determining the value of the feature weight is calculated based on the effect of the feature on the results of predictive analysis. The final result of determining the best features is applied to the data-set to improve the results of prediction accuracy.
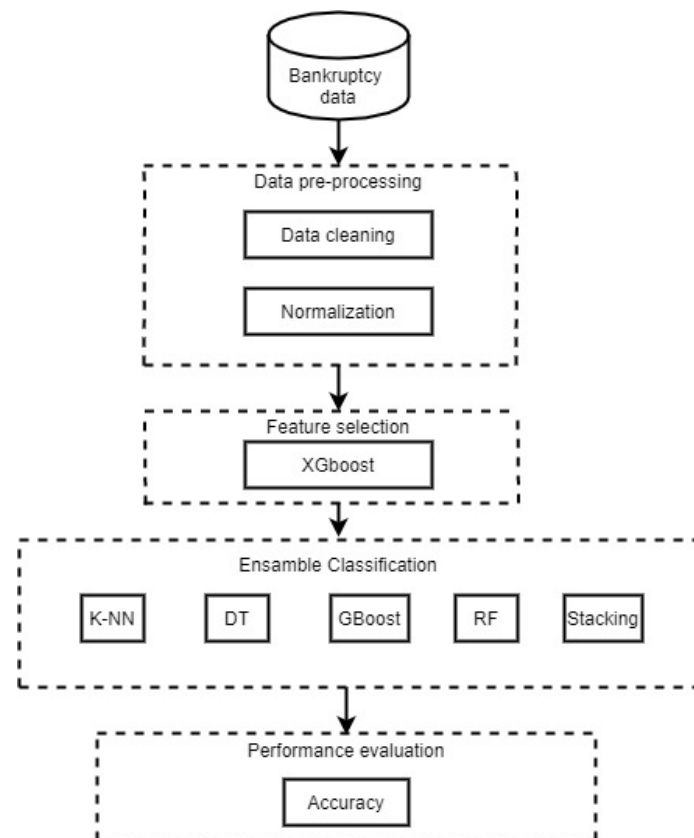

Figure 3. The research framework

The modeling process uses stacking ensemble learning, which is the process of combining several machine learning algorithms such as K-nearest neighbor, decision tree, gradient boosting tree and random forest [59]. Ensemble stacking is one of the ensemble learning methods and can use heterogeneous machine learning methods. Stacking ensemble learning uses meta-learning algorithms to find the best results for combining predictions from two or more basic machine learning algorithms. The stacking ensemble has the advantage of being able to take advantage of the work processes of several machine learning algorithm models that function well in classification or regression tasks and make predictions better than the work process of one machine learning model in ensemble learning.

4.   RESULTS AND DISCUSSION

The analysis process uses the google collab tool with the python programming language and the help of the scikit-learn, pandas, numpy libraries and other supporting libraries. The data-set used comes from Kaggle with a detailed data-set consisting of 5 CSV files which are combined into one to facilitate the prediction analysis process. The pre-processing data stage is scaling the data on the data-set using the standard scaler python library. The data scaling process was applied to each of the numerical data contained in the data-set. Data transformation is only performed on features used for the prediction process. This is because the target feature data are binary, namely 0 and 1, so there is no need for transformation through data scaling. The results of the scaling data are then analyzed at the feature importance stage.

4.1. Feature importance

The feature importance stage is the process of selecting the best features from the research dataset. The process of determining the best features uses an algorithm of feature importance from the XGBoost machine learning method. The important features are selected based on the weight value of each feature generated during the prediction analysis process. the best feature is selected based on the feature weight that is more than 10. Details of the best feature selection results are shown in Table 1.

Table 1. Fitur importance

Feature	Weight
Attribute 27	111
Attribute 34	81
Attribute 5	55
Attribute 46	41
Attribute 21	36
Attribute 35	25
Attribute 6	25
Attribute 58	23
Attribute 24	23
Attribute 56	20
Attribute 13	20
Attribute 41	17
Attribute 39	16
Attribute 22	16
Attribute 29	15
Attribute 47	14
Attribute 38	14
Attribute 26	13
Attribute 44	10
Attribute 30	10

The modeling stage is in the form of a normalized bankruptcy prediction analysis process through the data pre-processing stage. At this stage the dataset is analyzed using various machine learning methods. The prediction analysis process begins by dividing the data into training data and test data with a 75:25 ratio. The data sharing process was stratified and repeated. Stratified is a data sharing method based on the weight ratio ratio of the features for which the selected category. In this case the category feature selected is the target feature. Repeated is a method in which the data sharing process is repeated according to the parameters. The looping process is added with data shuffle, resulting in different data for each iteration. Cross-validation is included in this process to avoid overfit and underfit to maximize the quality of predictive analysis. Overfit is a model that is highly dependent on the dataset and has a high error value on the testing data. Underfit is a model that cannot fully understand the dataset being analyzed.

4.2. Stacking

The machine learning method used in the modeling stage of this study is stacking ensemble learning. Stacking means stacking, which means piling up the work process of machine learning methods to produce better predictive results. Machine learning methods that can be used in stacking can be selected heterogeneously. The type of stacking ensemble method used in this study is the classification of bankruptcy predictions.

The stacked machine learning algorithms in this study are K-nearest neighbor, decision tree, gradient boosting tree and random forest in this case called the base model. The base model can consist of many algorithms, but the more algorithms are used, the more resources and time it uses. Algorithms in the base model are not limited to just one model, they can also be used from many variations of the model according to research needs. This research process uses the classification method so that the algorithm used is

a type of classification. The result of the base model buildup is calculated by the meta learner. Meta learner is a machine learning algorithm that is used to analyze and combine the results of each base model in order to obtain a better prediction rate from the base model. The meta learner used in this study is LightGBM. The final result of the stacking model is a prediction generated by the meta learner. The accuracy details are shown in Figure 4. In Figure 4, it is shown that the difference between the models varies. The lowest level of accuracy is obtained in the decision tree algorithm with only 94.8%. The highest level of accuracy is obtained by the stacking model algorithm 97%.
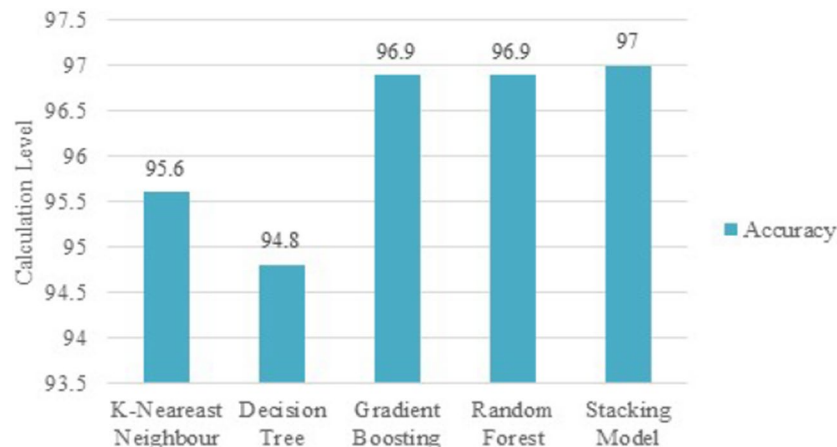

Figure 4. Detail accuracy

5.   CONCLUSION

In this study, a new method has been used to analyze bankruptcy predictions using the best feature selection and ensemble learning. The process of selecting the best features uses XGBoost's important features and the stacking method. The base model used is the K-nearest neighbor, decision tree, gradient boosting tree and random forest. The meta learner used is LightGBM. The stacking model accuracy results can outperform the base model accuracy with an accuracy rate of 97%.

REFERENCES

[1]   R. L. Constand and R. Yazdipour, "Firm failure prediction models: A critique and a review of recent developments," *Advances in Entrepreneurial Finance*, pp. 185-204, 2011, doi: 10.1007/978-1-4419-7527-0_10.
[2]   H. Li, D. Andina, and J. Sun, "Multiple proportion case-basing driven CBRE and its application in the evaluation of possible failure of firms," *International Journal of Systems Science*, vol. 44, no. 8, pp. 1409-1425, 2012, doi: 10.1080/00207721.2012.659686.
[3]   D. L. Olson, D. Delen, and Y. Meng, "Comparative analysis of data mining methods for bankruptcy prediction," *Decision Support Systems*, vol. 52, no. 2, pp. 464-473, 2012, doi: 10.1016/j.dss.2011.10.007.
[4]   L. Zhou, K. Keung Lai, and J. Yen, "Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimization," *International Journal of Systems Science*, vol. 45, no. 3, pp. 241-253, 2014, DOI: 10.1080/00207721.2012.720293.
[5]   L. Li and J. Sun, "Gaussian case-based reasoning for business failure prediction with empirical data in China," *Information Sciences*, vol. 3, no. 1-2, pp. 89-108, 2009, doi: 10.1016/j.ins.2008.09.003.
[6]   B. Prasetiyo, Alamsyah and M. A. Muslim, "Analysis of building energy efficiency dataset using naive bayes classification classifier," *Journal of Physics: Conference Series*, vol. 179, no. 3, 2019, Art. no. 032016, doi: 10.1088/1742-6596/1321/3/032016.
[7]   A. Nurzahputra, M. A. Muslim and B. Parsetiyo, "Optimization of C4.5 algorithm using meta-learning in diagnosing of chronic kidney diseases," *Journal of Physics: Conference Series*, vol. 179, no. 3, 2019, Art. no. 032022, doi: 10.1088/1742-6596/1321/3/032022.

[8]   Y. Ding, X. Song, and Y. Zen, "Forecasting financial condition of chinese listed companies based on support vector machine," *Expert Systems with Applications*, vol. 34, no. 4, pp. 3081-3089, 2008, doi: 10.1016/j.eswa.2007.06.037.

[9]   B. E. Edorgan, "Prediction of bankruptcy using support vector machines: An application to bank bankruptcy," *Journal of Statistical Computation & Simulation*, vol. 83, no. 8, pp. 1543-1555, 2013, doi: 10.1080/00949655.2012.666550.

[10] K. S. Shin, T. S. Lee, and H. -J. Kim, "An application of support vector machines in bankruptcy prediction model," *Journal of Financial Research,* vol. 28, no. 1, pp. 127-135, 2006, doi: 10.1016/j.eswa.2004.08.009.

[11] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405-417, 2017, doi: 10.1016/j.eswa.2017.04.006.

[12] R. Geng, I. Boseb, and X. Chen, "Prediction of financial distress: An empirical study of listed Chinese companies using data mining," *European Journal of Operational Research*, vol. 241, no. 1, pp. 236-247, 2015, doi: 10.1016/j.ejor.2014.08.016.

[13] S. Lessmann, B. Baesens, H.-V. Seow, L. C.Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124-136, 2015, doi: 10.1016/j.ejor.2015.05.030.

[14] C. Luo, Desheng Wu, and Dexiang Wu, "A deep learning approach for credit scoring using credit default swaps," *Engineering Applied Artificial Intelligence*, vol. 65, pp. 465-470, 2017, doi: 10.1016/j.engappai.2016.12.002.

[15] M. Stankova and D. Hampel, "Bankruptcy prediction of engineering companies in the EU using classification methods," *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*, vol. 66, no. 5, pp. 1347-1355, 2018, doi: 10.11118/actaun201866051347.

[16] S. K. Shrivastav and P. J. Ramudu, "Bankruptcy prediction and stress quantification using support vector machine: evidence from Indian Banks," *Risks*, vol. 8, no. 2, pp. 1-22, 2020, doi: 10.3390/risks8020052.

[17] A. Gepp, K. Kumar, and S. Bhattacharya, "Business failure prediction using decision trees," *journal of forecasting*, vol. 29, no. 6, pp. 536-555, 2010, doi: 10.1002/for.1153.

[18] D. Delen, C. Kuzey, and A. Uyar, "Measuring firm performance using financial ratios: A decision tree approach," *Expert Systems with Applications*, vol. 40, no. 10, pp. 3970-3983, 2013, doi: 10.1016/j.eswa.2013.01.012.

[19] S. Y. Kim and A. Upneja, "Predicting restaurant financial distress using decision tree and adaboosted decision tree models," *Economic Modelling*, vol. 36, pp. 354-362, 2014, doi: 10.1016/j.econmod.2013.10.005.

[20] N. Ocal, M. K. Ercan, and E. Kadioglu, "Predicting financial failure using decision tree algorithms: An empirical test on the manufacturing industry at Borsa Istanbul," *International Journal of Economics and Finance*, vol. 7, no. 7, pp. 189-206, 2015, DOI: 10.5539/ijef.v7n7p189.

[21] A. Gepp and K. Kumar, "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques," *Procedia Computer Science*, vol. 54, pp. 396-404, 2015, doi: 10.1016/j.procs.2015.06.046.

[22] H. Sherma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *International Journal Science Research*, vol. 5, no. 4, pp. 2094-2097, 2016.

[23] S. H. Syed Nor, S. Ismail, and B. W. Yap Yap, "Personal bankruptcy prediction using decision tree model," *Journal of Economics Finance and Administrative Science*, vol. 4, no. 7, pp. 157-170, 2020, doi:10.1108/JEFAS-08-2018-0076.

[24] R. L. Wilson and R. Sharda, "Bankruptcy prediction using neural networks," *Decision Support System*, vol. 11, no. 5, pp. 545-557, 1994, doi: 10.1016/0167-9236(94)90024-8.

[25] A. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," *IEEE Transactions on Neural Networks,* vol. 12, no. 4, pp. 929-935, 2001, doi: 10.1109/72.935101.

[26] M. Anandarajan, P. Lee, and A. Anandarajan, "Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks," *International Journal Intellegent System Account*, vol. 10, no. 2, pp. 69-81, 2001, doi: 10.1002/isaf.199.

[27] M. J. Kim and D. K. Kang, "Ensemble with neural networks for bankruptcy prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3373-3379, 2010, doi: 10.1016/j.eswa.2009.10.012.

[28] A. A. Kasgari, M. Divsalar, M. R. Javid, and S. J. Ebrahimian, "Prediction of bankruptcy Iranian corporations through artificial neural network and probit-based analyses," *Neural Computing Application*, vol. 23, no. 3-4, pp. 927-936, 2013, doi: 10.1007/s00521-012-1017-z.

[29] Z. Dong *et al.* "An effective computational model for bankruptcy prediction using kernel extreme learning machine approach," *Computational Economics*, vol. 49, no. 2, pp. 325-341, 2017, doi: 10.1007/s10614-016-9562-7.

[30] G. P. Naidu and K. Govinda, "Bankruptcy Prediction Using Neural Networks," *2018 2nd International Conference on Inventive Systems and Control (ICISC),* 2018, pp. 248-251, doi: 10.1109/ICISC.2018.8399072.

[31] T. Hosaka, "Bankruptcy prediction using imaged financial ratios and convolutional neural networks," *Expert Systems with Applications,* vol. 117, pp. 287-299, 2019, doi: 10.1016/j.eswa.2018.09.039.

[32] D. Liang, C.-C. Lu, C.-Fong Tsai, G.-An Shiha, "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," *European Journal of Operational Research*, vol. 252, no. 2, pp. 561-572, 2016., doi: 10.1016/j.ejor.2016.01.012.

[33] H. A. Alaka *et al.,* "Systematic review of bankruptcy prediction models: Towards a framework for tool selection," *Expert Systems With Applications,* vol. 94, pp. 164-184, 2018, doi: 10.1016/j.eswa.2017.10.040.

[34] Y. Shi and X. Li, "An overview of bankruptcy prediction models for corporate firms: A systematic literature review," *Intangible Capital,* vol. 15, no. 2, pp. 114-127, 2019, doi: 10.3926/ic.1354.

[35] Y. Qu, P. Quan, M. Lei, and Y. Shi, "Review of bankruptcy prediction using machine learning and deep learning techniques," *Procedia Computer Science*, vol. 162, pp. 895-899, 2019, doi: 10.1016/j.procs.2019.12.065.

[36] S. S. Devi and Y. Radhika, "A survey on machine learning and statistical techniques in Bankruptcy Prediction," *International Journal of Machine Learning and Computing*, vol. 8, no. 2, pp. 133-139, 2018, doi: 10.18178/ijmlc.2018.8.2.676.

[37] P. Gnip and P. Drotar, "Ensemble methods for strongly imbalanced data: bankruptcy prediction," *2019 IEEE 17th International Symposium on Intelligent Systems and Informatics (SISY),* 2019, pp. 155-160, doi: 10.1109/SISY47553.2019.9111557.

[38] G. Wang, J. Ma, and S. Yang, "An improved boosting based on feature selection for corporate bankruptcy prediction," *Expert Systems With Applications*, vol. 41, no. 5, pp. 2353-2361, 2014, doi: 10.1016/j.eswa.2013.09.033.

[39] N. Wang, "Bankruptcy prediction using machine learning," *Journal of Mathematical Finance*, vol. 7, no. 4, pp. 908-918, 2017, doi: 10.4236/jmf.2017.74049.

[40] H. Faris, W. Manaseer, M. Saadeh, A. Mora, P. A. Castillo, I. Aljarah, "Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market," *Progress in Artificial Intelligence*, vol. 9, no. 1, pp. 31-53, 2019.

[41] Z. Chen, W. Chen, and Y. Shi, "Ensemble learning with label proportions for bankruptcy prediction," *Expert Systems With Applications,* vol. 146, 2020, Art. no. 113155, doi: 10.1016/j.eswa.2019.113155.

[42] T. Le Minh, T. Vo, B. Vo, Mi Y. Lee, and S. Wook Baik, "A hybrid approach using oversampling technique and cost-sensitive learning for bankruptcy prediction," *Complexity,* vol. 2019, no. 2, pp. 1-12, 2019, Art. no. 8460934, doi: 10.1155/2019/8460934.

[43] S. Vellemcheti and P. Singh "Class imbalance deep learning for bankruptcy prediction," *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, 2020, pp. 421-425, doi: 10.1109/ICPC2T48082.2020.9071460.

[44] S. Smiti and M. Soui, "Bankruptcy prediction using deep learning approach based on Borderline SMOTE," *Information Systems Frontiers*, vol. 22, no. 5, pp. 1067-1083, 2020.

[45] T. H. Kanga, S. D. James, and F. Fabian, "Real options and strategic bankruptcy," *Journal of Business Research,* vol. 117, pp. 152-162, 2020, doi: 10.1016/j.jbusres.2020.05.057.

[46] W. C. Lin, Y.-H. Lu, and C.-F. Tsai, "Feature selection in single and ensemble learning-based bankruptcy prediction models," *Expert Systems,* vol. 36, no. 1, pp. 1-8, 2018, doi: 10.1111/exsy.12335.

[47] H. Son, C. Hyun, D. Phan, H. J. Hwang, "Data analytic approach for bankruptcy prediction," *Expert Systems With Applications*, vol. 138, 2019, Art. no. 112816, doi: 10.1016/j.eswa.2019.07.033.

[48] J. Nobre and R. F. Neves, "Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to trade in the financial markets," Expert Systems With Applications, vol. 125, pp. 181-194, 2019, doi: 10.1016/j.eswa.2019.01.083.

[49] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.

[50] Y. Freund, and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society For Artificial Intelligence*, vol. 14, no. 5, pp. 771-790, 1999.

[51] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241-259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[52] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, no. 1, pp. 49-64, 1996.

[53] M. J. Van der Laan, E. C. Polley, A. E. Hubbard, "Super learner," *Statistical Applications in Genetics and Molecular*, vol. 6, no. 1, 2007, Art. no. 25, doi: 10.2202/1544-6115.1309.

[54] M. Zieba, S. K. Tomczak, J. M. Tomczak, "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction," *Expert Systems With Applications*, vol. 58, pp. 93-101, 2016, doi: 10.1016/j.eswa.2016.04.001.

[55] T. Mustaqim, K. Umam and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *Journal of Physics: Conference Series*, vol. 1567, pp. 8-15, 2020, Art. no. 032024, doi: 10.1088/1742-6596/1567/3/032024.

[56] I. Ashari, M. A. Muslim, and A. Alamsyah, "Comparison Performance of Genetic Algorithm and Ant Colony Optimization in Course Scheduling Optimizing," *Scientific Journal Informatics*, vol. 3, no. 2, pp. 149-158, 2016, doi: 10.15294/sji.v3i2.7911.

[57] A. R. Safitri and M. A. Muslim "Improved Accuracy of Naive Bayes Classifier for Determination of Customer Churn Uses SMOTE and Genetic Algorithms," *Journal Soft Computing Exploration*, vol. 1, no. 1, pp. 70-75, 2020, doi: 10.52465/joscex.v1i1.5.

[58] Y. Guo, " Credit risk assessment of P2P lending platform towards big data based on BP neural network*," Journal of Visual Communication and Image Representation*, vol. 71, 2020, Art. no. 102730, doi: 10.1016/j.jvcir.2019.102730.

[59] D. Liang, C.-F. Tsai, H.-Y. (Richard) Lu, L.-S. Chang, "Combining corporate governance indicators with stacking ensembles for financial distress prediction," *Journal of Business Research*, vol. 120, pp. 137-146, 2020, doi: 10.1016/j.jbusres.2020.07.052.

BIOGRAPHIES OF AUTHORS

Much Aziz Muslim PhD candidate in the faculty of management technology at Universiti Tun Hussein Onn Malaysia (UTHM). The scope of research he is currently working on is in the fields of Computer Science, Data Mining and Networking. besides that he is also a lecturer in the computer science department of the Universitas Negeri Semarang (UNNES).
Email: a212muslim@mail.unnes.ac.id

Yosza Dasril received PhD and Master Degree degree in Applied Mathematics from Univeriti Putra Malaysia in 2003 and 1999 respectively and Bachelor Degree in Mathematics from Universitas Riau, Indonesia in 1994. From 1999 to 2006, he was a lecturer in the Universiti Malaysia Terengganu (UMT). In the middle of 2006 into Nov 2019, he serve as a Senior Lecturer at the Department of Electronics and Computer Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM). Currently, he is a Lecturer at Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia (UTHM). His research interests are in Optimization, Engineering Mathematics and Computational Mathematics.
Email: yosza@utem.edu.my