Combination Of Weak Learner And Strong On Stacking To Increase Bankruptcy Risk Prediction

Wiena Faqih Abror[1], Much Aziz Muslim[2], Dwika Ananda Agustina Pratiwi[3]
[1] Universitas Negeri Semarang, Indonesia
[2] Universiti Tun Hussein Onn Malaysia, Malaysia
* Corresponding Author

ARTICLE INFO

Article history:

Received July 28, 2021
Revised August 28, 2021
Accepted September 28, 2021

Keywords:

Machine Learning;
Prediction;
Stacking.

ABSTRACT

Bankruptcy is a fatal phenomenon for agencies. Minimizing the risk of bankruptcy can be done through prediction techniques using machine learning. Machine learning algorithms such as k-nearest neighbors can be used as modeling to generate accuracy values. Although in machine algorithms there are categories that distinguish between weak learners and strong learners. Both categories can be combined using the stacking ensemble method to improve accuracy performance. Such as using the k-nearest neighbors algorithm, decision tree, naïve Bayes, gradient boosting decision tree, light gradient boosting machine, and extreme gradient boosting machine. The results of research using this algorithm have an accuracy performance of 99.23%.

Cite Article:

Wiena Faqih Abror, Much Aziz Muslim, Dwika Ananda Agustina Pratiwi, "Combination Of Weak Learner And Strong On Stacking To Increase Bankruptcy Risk Prediction" *Indonesian Community on Optimization and Computer Application*, vol. 1, no. 1, pp. xx-xx, 2022, doi: .

1. INTRODUCTION

Bankruptcy is a fatal phenomenon that every agency avoids. The effort to minimize bankruptcy is to predict bankruptcy based on agency financial reports [1]bankr. Machine learning algorithms can help with the bankruptcy prediction process [2], [3]. Machine learning algorithms that can be used to predict bankruptcy are k-nearest neighbors [4]–[6], naïve Bayes [7], [8], decision trees [9], [10], gradient boosting decision trees [11], [12], light gradient boosting machines [13], [14]. The type of method used in predicting bankruptcy is classification. Classify agencies labeled as bankrupt and agencies labeled as not bankrupt [15].

There are different categories in the use of algorithms in machine learning [16]. There are two different categories regarding algorithms, namely weak learners and strong learners. The weak concept in machine learning makes the naïve method a reference, besides that random guessing in the formation of predictions is a special characteristic of machine learning algorithms called a weak learner [16], [17]. Examples of weak learner algorithms are naïve Bayes, decision trees, and k-nearest neighbors [18].

Strong learner algorithms have arbitrary performance characteristics [19]. The authority possessed by algorithms with strong learner characteristics can be regulated through hyperparameters which can directly improve prediction accuracy [20]. Strong learner algorithms can also be built through weak learner algorithms. Examples of strong learner algorithms are gradient boosting decision trees, light gradient boosting machines, and extreme gradient boosting [18].

Increasing the results of accuracy in making predictions can be built through a combination of weak learner algorithms and strong learner algorithms [21]. The stacking method is a solution in combining the two categories. The mechanism used by the stacking method will carry out learning through two stages [1],

[22]. The first stage of the process is grouped in the base learner container, using the original dataset and then trained through the algorithm stacked in the base learner container to produce a new dataset. In the second stage of the process, the newly formed dataset will be retrained through the meta learner container [23]. The end result of the meta learner will result in the performance of the dataset that has been trained [24].

Lin [8] in his research used a Taiwanese dataset using decision trees, naïve Bayes, and k-nearest neighbors to find error values. It was found that the decision tree has a lower error rate than naïve Bayes and k-nearest neighbors. Wyrobek [25] uses the tree that has been formed to improve the next tree in the learning process using a gradient boosting decision tree on the polish dataset. Muslim [23] uses the light gradient boosting machine algorithm in the stacking method as a meta learner on the polish dataset. In this research, we try to combine the algorithms used to get the best model for predicting the Taiwanese dataset.

2. METHODS

The research method was built through three stages, data preprocessing, ensemble stacking, and performance evaluation. Referring to the characteristics of the dataset used, the Taiwanese dataset has imbalance properties. Imbalanced datasets can have an impact on performance evaluation because the accuracy results obtained are not optimal. Imbalanced datasets will be balanced using the synthetic minority over-sampling technique (SMOTE) [26]–[28]. SMOTE will rebuild the dataset using the nearest neighbor modeling. So the dataset that is formed still has a correlation with the existing dataset [28], [29]. The research flow is in Figure 1.

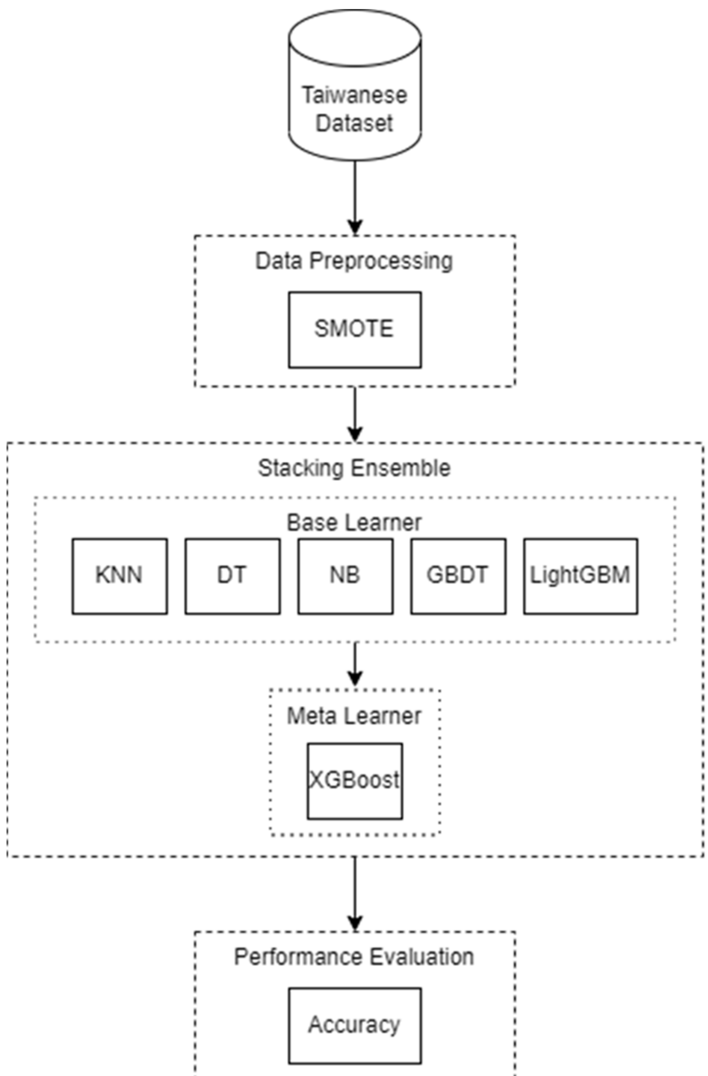
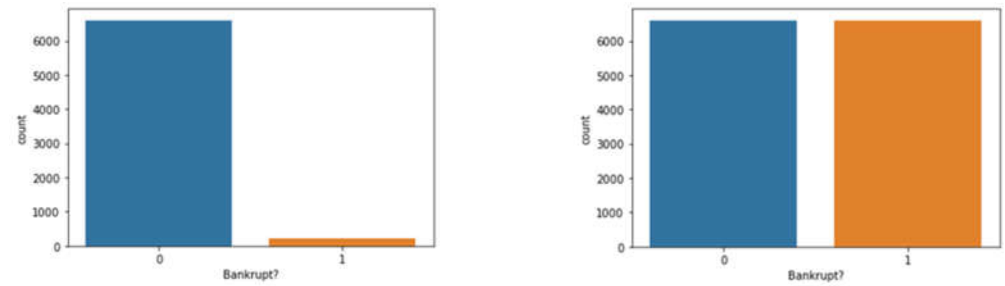


Fig. 1. Proposed Method

The balanced dataset is then modeled using the stacking ensemble method. Data modeling using the stacking ensemble method will be formed using k-nearest neighbors, decision trees, naïve bayes, gradient boosting decision trees, and light gradient boosting machines which at the stacking stage are part of the base learner. The modeled dataset forms a new array to be modeled again in the last stage of stacking, namely the meta learner using extreme gradient boosting. The modeling results from stacking are evaluated using a confusion matrix through k-fold cross validation to assess the accuracy of the model created [5].

3. RESULTS AND DISCUSSION

The comparison of the classification labels that have been calculated is 6599 data for label 0 which means not bankrupt and 220 data for label 1 which means bankrupt. After the dataset is balanced, it will show that the two labels show the same amount of data, namely 6599 data. As shown in Figure 2.



(a) (b)

Fig. 2. (a) Dataset imbalanced; (b) Dataset balanced

Data that has been balanced beforehand is modeled using a single classifier. Modeling using a single classifier aims to show the accuracy performance of the dataset that is modeled using the algorithm used in stacking. The algorithms used are k-nearest neighbors, decision tree, naïve bayes, gradient boosting decision tree, and light gradient boosting machine. The model performance results are visualized in Figure 3.

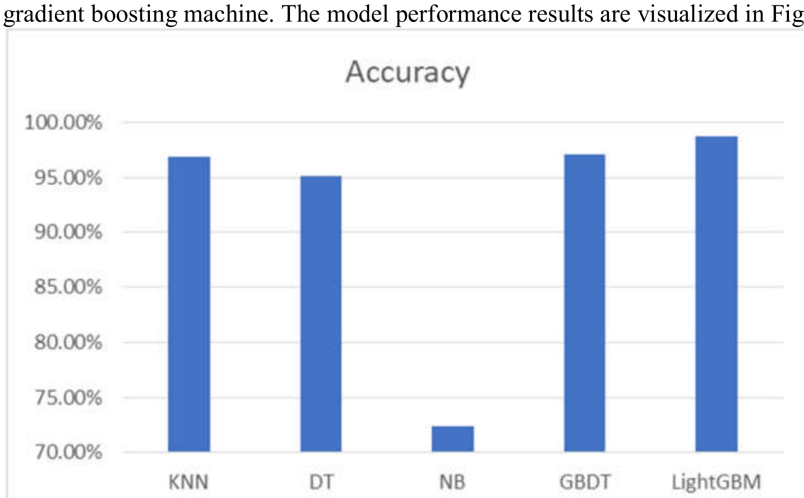


Fig. 3. Comparison single algorithms

The first modeling using k-nearest neighbors produces an accuracy performance of 96.87%. The second model uses the decision tree algorithm to provide an accuracy of 95.14%. The third model using naïve Bayes produces an accuracy of 72.34%. The fourth model uses a gradient boosting decision tree to produce an accuracy of 97.13%. The fifth model uses the light gradient boosting machine algorithm to provide an accuracy of 98.74%.

The single modeling for each classifier used is intended to compare the modeling results using the stacking ensemble method. Modeling using the stacking ensemble will go through the base learner stages which stack the k-nearest neighbor algorithm, decision tree, naïve bayes, gradient boosting decision tree, and light gradient boosting machine to produce a new dataset array. The new dataset array from base learner will be remodeled using extreme gradient boosting. The comparison of the algorithms used is shown in Figure 4.

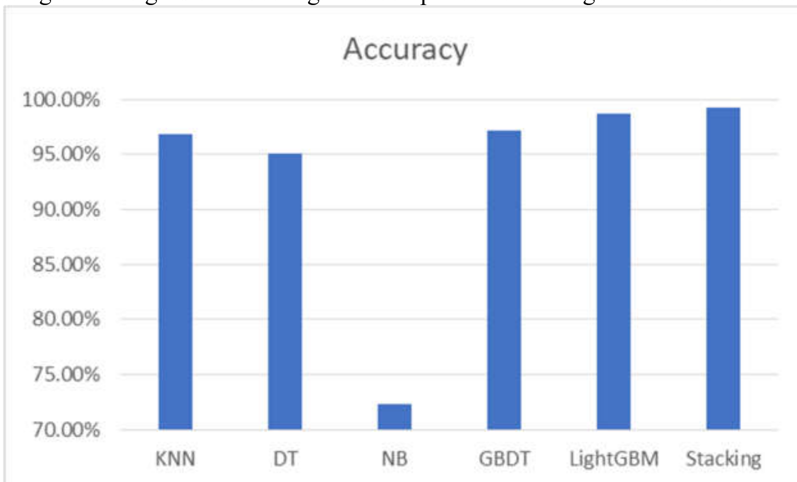


Fig. 4. Comparison single algorithms and Stacking Ensemble

Based on the comparison of the single algorithm used modeling with the stacking ensemble technique shows the difference. Modeling using the stacking ensemble technique produces an accuracy performance of 99.23%. The accuracy performance results show that modeling using the stacking ensemble technique is the best among the single models used

4. CONCLUSION

Minimizing the phenomenon of bankruptcy can use prediction efforts using machine learning algorithms. The research was conducted using k-nearest neighbors, decision trees, naïve bayes, gradient boosting decision trees, and light gradient boosting machines which were stacked using the stacking ensemble method. The stacking ensemble method applied uses extreme gradient boosting as a meta learner for the final modeling. The modeling that was created produces an accuracy of 99.23% on the Taiwanese Bankruptcy Dataset

DECLARATION

Author Contribution

The research method was built through three stages, data preprocessing, ensemble stacking, and performance evaluation.

Funding

Please add: “This research received no external funding” or “This research was funded by NAME OF FUNDER, grant number XXX” and “The APC was funded by XXX”.

Acknowledgments

In this section, you can acknowledge any support given which is not covered by the author con-tribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflict of Interest

Declare conflicts of interest or state “The authors declare no conflict of interest.”

REFERENCES

[1] D. Liang, C.-F. Tsai, H.-Y. (Richard) Lu, and L.-S. Chang, “Combining corporate governance indicators with stacking ensembles for financial distress prediction,” J. Bus. Res., vol. 120, pp. 137–146, 2020, doi: https://doi.org/10.1016/j.jbusres.2020.07.052.

[2] R. C. Lacher, P. K. Coats, S. C. Sharma, and L. F. Fant, “A neural network for classifying the financial health of a firm,” Eur. J. Oper. Res., vol. 85, no. 1, pp. 53–65, 1995, doi: https://doi.org/10.1016/0377-2217(93)E0274-2.

[3] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” Comput. Electr. Eng., vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/J.COMPELECENG.2013.11.024.
[4] W. Xing and Y. Bei, “Medical Health Big Data Classification Based on KNN Classification Algorithm,” IEEE Access, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.
[5] A. Mustaqeem, S. M. Anwar, M. Majid, and A. R. Khan, “Wrapper method for feature selection to classify cardiac arrhythmia,” Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp. 3656–3659, Sep. 2017, doi: 10.1109/EMBC.2017.8037650.
[6] A. Susanto, D. Sinaga, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, “A High Performace of Local Binary Pattern on Classify Javanese Character Classification,” Sci. J. Informatics, vol. 5, no. 1, p. 8, 2018, doi: 10.15294/sji.v5i1.14017.
[7] H. Wang and X. Liu, “Undersampling bankruptcy prediction: Taiwan bankruptcy data,” PLoS One, vol. 16, no. 7, p. e0254030, Jul. 2021, [Online]. Available: https://doi.org/10.1371/journal.pone.0254030
[8] W.-C. Lin, Y.-H. Lu, and C.-F. Tsai, “Feature selection in single and ensemble learning-based bankruptcy prediction models,” Expert Syst., vol. 36, no. 1, p. e12335, 2019, doi: https://doi.org/10.1111/exsy.12335.
[9] C.-F. Tsai, Y.-F. Hsu, and D. C. Yen, “A comparative study of classifier ensembles for bankruptcy prediction,” Appl. Soft Comput., vol. 24, pp. 977–984, 2014, doi: https://doi.org/10.1016/j.asoc.2014.08.047.
[10] R. G. Mantovani, A. L. D. Rossi, E. Alcobaça, J. Vanschoren, and A. C. P. L. F. de Carvalho, “A meta-learning recommender system for hyperparameter tuning: Predicting when tuning improves SVM classifiers,” Inf. Sci. (Ny)., vol. 501, pp. 193–221, 2019, doi: https://doi.org/10.1016/j.ins.2019.06.005.
[11] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” Front. Neurorobot., vol. 7, no. DEC, p. 21, 2013, doi: 10.3389/FNBOT.2013.00021/BIBTEX.
[12] J. Yang, C. Zhao, H. Yu, and H. Chen, “Use GBDT to Predict the Stock Market,” Procedia Comput. Sci., vol. 174, pp. 161–171, 2020, doi: https://doi.org/10.1016/j.procs.2020.06.071.
[13] G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in Advances in Neural Information Processing Systems, 2017, vol. 30. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
[14] M. A. Muslim, Y. Dasril, M. Sam’an, and Y. N. Ifriza, “An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending,” Indones. J. Electr. Eng. Comput. Sci., vol. 28, no. 2, pp. 1002–1011, 2022.
[15] D. Liang, C.-C. Lu, C.-F. Tsai, and G.-A. Shih, “Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study,” Eur. J. Oper. Res., vol. 252, no. 2, pp. 561–572, 2016, doi: https://doi.org/10.1016/j.ejor.2016.01.012.
[16] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, “A survey on ensemble learning,” Front. Comput. Sci., vol. 14, no. 2, pp. 241–258, 2020, doi: 10.1007/s11704-019-8208-z.
[17] Y. Freund, “Boosting a weak learning algorithm by majority,” Inf. Comput., vol. 121, no. 2, pp. 256–285, 1995.
[18] R. Saravanan and P. Sujatha, “A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification,” in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 945–949.
[19] G. Cho, J. Yim, Y. Choi, J. Ko, and S.-H. Lee, “Review of machine learning algorithms for diagnosing mental illness,” Psychiatry Investig., vol. 16, no. 4, p. 262, 2019.
[20] B. Mahesh, “Machine learning algorithms-a review,” Int. J. Sci. Res. (IJSR).[Internet], vol. 9, pp. 381–386, 2020.
[21] D.-C. Feng et al., “Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach,” Constr. Build. Mater., vol. 230, p. 117000, 2020.
[22] J. Dou et al., “Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan,” Landslides, vol. 17, no. 3, pp. 641–658, 2020, doi: 10.1007/s10346-019-01286-5.
[23] M. A. Muslim and Y. Dasril, “Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning.,” Int. J. Electr. Comput. Eng., vol. 11, no. 6, 2021.
[24] M. A. Muslim and B. Prasetiyo, “Implementation twofish algorithm for data security in a communication network using library chilkat encryption activex,” J. Theor. Appl. Inf. Technol., vol. 84, no. 3, p. 370, 2016.
[25] J. Wyrobek and K. Kluza, “Efficiency of gradient boosting decision trees technique in Polish companies’ bankruptcy prediction,” in International Conference on Information Systems Architecture and Technology, 2018, pp. 24–35.
[26] A. Fernández, S. Garcia, F. Herrera, and N. V Chawla, “SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary,” J. Artif. Intell. Res., vol. 61, pp. 863–905, 2018.
[27] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
[28] B. Prasetiyo, Alamsyah, M. A. Muslim, and N. Baroroh, “Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique,” J. Phys. Conf. Ser., vol. 1918, no. 4, p. 42002, 2021, doi: 10.1088/1742-6596/1918/4/042002.
[29] A. R. Safitri and M. A. Muslim, “Improved accuracy of naive bayes classifier for determination of customer churn uses smote and genetic algorithms,” J. Soft Comput. Explor., vol. 1, no. 1, pp. 70–75, 2020.

[30]