Optimization Analysis of Neural Network Algorithms Using Bagging Techniques on Classification of Date Fruit Types

Rully Pramudita
Informatic Management
Bina Insani University
Bekasi, Indonesia
rullypramudita@binainsani.ac.id

Solikin
Information System
Bina Insani University
Bekasi, Indonesia
solikin@binainsani.ac.id

Nadya Safitri
Software Engineering
Bina Insani University
Bekasi, Indonesia
nadyasafitri@binainsani.ac.id

2022 Seventh International Conference on Informatics and Computing (ICIC) | 979-8-3503-4571-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIC56845.2022.10006986

***Abstract*— To perform an optimization One of the machine learning algorithms is by using Ensemble Learning. The main problem in this study is to increase the accuracy of the neural network algorithm in classifying data. This study aims to optimize the neural network algorithm in classifying. There are many optimization algorithms in classifying. In this study, we tried to optimize using the bagging technique for classification. This study uses types of dates as a case study. The results of the model evaluation show that the neural network algorithm plus the bagging technique produces an accuracy rate of 96%, and this has a higher accuracy of 1% compared to the neural network algorithm alone.**

***Keywords*— Algorithms, Bagging, Classification, Optimization, Neural Network.**

I. INTRODUCTION

Accuracy value of a classification algorithm shows whether or not an algorithm is good. An algorithm the accuracy value can be increased again with some other technique or algorithm, for example Using Bootstrap Aggregating (Bagging) or Adaptive Boosting (AdaBoost) [1].

The Bagging Algorithm provides better stability for the model in his research [2]. Accuracy Results and F1 models formed by the Bagging method tend to show better value performance than the AdaBoost method on evaluation metrics [3]. Bootstrap is one resampling techniques are important for estimating the mean and variance, among other properties, of statistical estimator [4]. The Aggregating bootstrap method is proven to be effective in reducing bias in the neural network system which results in increased precision prediction results [5]. In Ryno du Plooy's research, he said that the bagging method is more accurate than ANN in guaranteeing a zero price [6]. Based on the experiments conducted, it shows that using the bagging method produces a better level of predictive efficiency than using a single learning algorithm [7].

Base on Green Arther Sandag research, bagging classifier algorithm with 0.13% more accurate results than using the XGBoost algorithm [8]. Based Laksana research, it says that the calculation of accuracy, RMSE and classification error, the C4.5 Bagging algorithm is more suitable than the C4.5 method, the accuracy calculation value is 46%, the Classification Error is 54, and the RMSE is 0.690, so the C4.5 Bagging algorithm is better compared to the C4.5 algorithm with an accuracy value of 46%. This shows that the bagging method can increase the level of accuracy of the algorithm [9]. By applying the bagging technique for ensemble-based classification on the naïve Bayes algorithm can increase the accuracy of 3.00%. With an initial accuracy of 77.00%, after applying the technique bagging to 80.00% [10]. Based on research from Hong-Jun Yoon, it resulted that the partitioning and bagging strategies succeeded in producing higher performance scores [11]. From Nursimpati's research, the results of calculating performance data using the bagging technique can increase accuracy by 2.381% compared to the naïve Bayes classification algorithm [12]. Results show that the BS technique is superior and more informative for the electric grid operators than a technique based on the use of the quantiles of the ensemble model predictions [13].

From the research results, with applying the bagging technique to ensemble-based classification on the C4.5 algorithm can increase accuracy by 12.86% [14]. The results of Abdollahi's research on the use of neural network-based ensemble learning methods to diagnose and predict chronic diseases show high performance, achieving accuracies of 98.5, 99, and 100% respectively [15]. The bagging method shows good performance in terms of precision and recall, an increase of 0.98% in his research [16]. Based on research conducted by Possebon, the bagging method produces the best performance in the form of an accuracy of 99.97% and a false positive value of 0.00018% [17]. The results of this study say that bagging and neural network techniques improve accuracy by 1-5 percent [18]. Bagging and boosting ensembles can reduce more than 30% in cases of PM10 and 20% in cases of PM2.5 [19]. Of all the experimental results of machine learning algorithms and ensemble techniques that are compared based on accuracy, sensitivity, specificity and f1 score, bagging is the best method [20].

There are many previous studies that discuss classification using neural network algorithms and optimization techniques. Based on previous studies, then this research will discuss the optimization of neural network algorithms in classifying. The optimization technique used is bagging. The results and discussion section will discuss the level of accuracy of the neural network algorithm and its optimization results, the dataset that will be used is public data from Kaggle.

II. METHODS

The stages in conducting this research start from the data preprocessing process in which there is a data source preparation process and is divided into training data and data testing. Furthermore, after the data is ready, the next stage is modeling which is the process of determining the model to be used, in this study the model used is the neural network algorithm and the bagging optimization algorithm.

In this stage, the learning process will be carried out according to the specified classification model as well as the optimization process based on training data and testing data. The results of the modeling process will be used to carry out the last stage, namely evaluation which is an assessment of the accuracy of the selected model and compares it with the results of the optimization carried out. Figure 1 explains the flowchart of the stages of the research carried out.

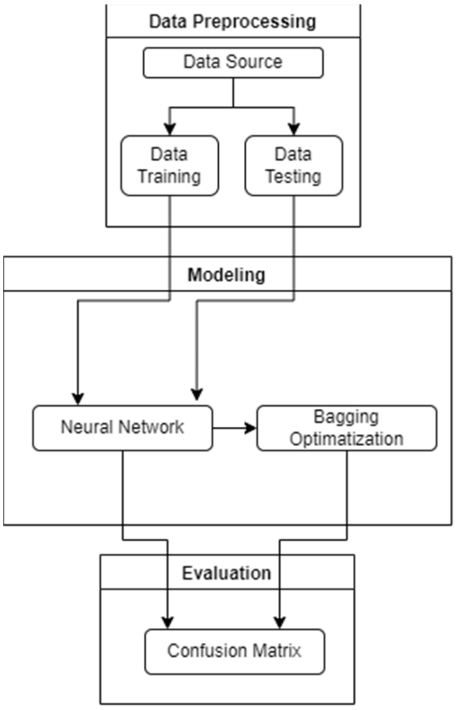

Fig. 1. Methodology

III. RESULT

A. Data Preprocessing

The dataset obtained at this stage will be carried out in the data preparation process, and will be used as training data and testing data. Training data and testing data will then be used as input for the classification process by the neural network algorithm and bagging optimization. From a total dataset of 898 records, 100 records will be used as testing data taken by sampling from the total dataset.

B. Data Modeling

At this stage, the process of making a model and carrying out the data training process is carried out using a predetermined classification model. After the model is trained using the training data, then the next step is to test the model using data testing. This process is assisted by using Rapid Miner tools. In the first model it is carried out using a neural network algorithm, the following figure 2 explains the neural network classification model.

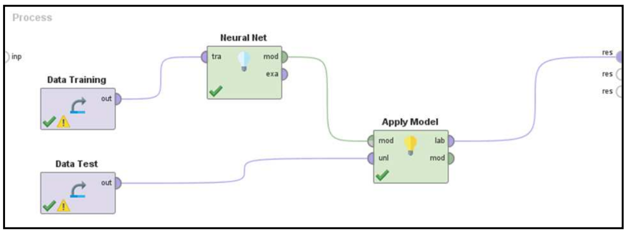

Fig. 2. Neural Network Classification Models

In Figure 2 above it can be seen that the neural network classification model gets input in the form of training data of 898 types of dates and all their attributes. And produces output in the form of a model that has been trained, this model becomes the next input, namely the testing process in applying the model to be tested using 100 data testing data. The results of this first model are in the form of an example set table which contains the prediction column of the target label and the confidence value of each attribute. Table 1 describes a sample confidence value from several class prediction results.

TABLE I. SAMPLE DISCREPANCIES IN CONFIDENCE MODEL NN VALUES

Class	confidence(BERHI)	confidence(DEGLET)	confidence(DOKOL)	confidence(IRAQI)	confidence(ROTANA)	confidence(SAFAVI)	confidence(SOGAY)	prediction(Class)	Information
DEGLET	0,000	0,407	0,001	0,000	0,129	0,005	0,458	SOGAY	Not accordance
DOKOL	0,002	0,119	0,417	0,000	0,000	0,004	0,458	SOGAY	Not accordance
SAFAVI	0,000	0,000	0,529	0,000	0,000	0,471	0,000	DOKOL	Not accordance
SOGAY	0,001	0,104	0,000	0,000	0,847	0,000	0,048	ROTANA	Not accordance
SOGAY	0,001	0,757	0,002	0,000	0,019	0,001	0,221	DEGLET	Not accordance

Based on table 1, it is known that there are several confidence values that have a value close to 1, not in accordance with the targeted class or label. For example, the DEGLET date type has the greatest confidence value in the SOGAY date column so that the prediction results are not appropriate. Out of a total of 100 testing data, there are 5 data examples that produce inappropriate predictions. Next is to do the second modeling by adding a bagging technique to optimize the neural network algorithm. Figure 3 describes the shape of the second model.

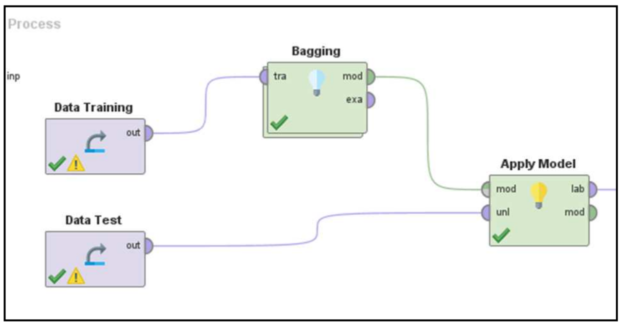

Fig. 3. Neural network classification optimization model using bagging

Figure 3 above shows the neural network algorithm optimization model using the bagging technique, this bagging process has the same input as the first model, namely 898 training data, after which it produces an output in the form of a training model which will be tested using data testing through the apply model. This model has used bagging techniques to carry out optimization, this is done to see whether better accuracy results will be obtained than the first model or the same or even lower accuracy. After the process of applying the model, it will return to the example set table as described in table 2.

TABLE II. SAMPLE DISCREPANCY VALUE CONFIDENCE MODEL NEURAL NETWORK OPTIMIZATION

Class	confidence(BERHI)	confidence(DEGLET)	confidence(DOKOL)	confidence(IRAQI)	confidence(ROTANA)	confidence(SAFAVI)	confidence(SOGAY)	prediction(Class)	Information
DEGLET	0,000	0,325	0,000	0,000	0,194	0,009	0,470	SOGAY	Not accordance
DOKOL	0,002	0,277	0,322	0,000	0,000	0,003	0,396	SOGAY	Not accordance
SOGAY	0,002	0,153	0,000	0,000	0,771	0,000	0,074	ROTANA	Not accordance
SOGAY	0,001	0,853	0,004	0,000	0,013	0,001	0,128	DEGLET	Not accordance

Based on table 2 above it can be seen that the results of the second modeling using the optimization of the neural network algorithm total 4 data, the number of prediction discrepancies seen from the confidence value which is close to 1 but not on target. With this the modeling process has been completed, next is the evaluation stage.

C. Evaluation

The evaluation is carried out to determine the level of accuracy of the model that has been formed, the accuracy value is obtained using a classification performance measurement method that produces a confusion matrix. To get the results of the confusion matrix, the step that needs to be done is to add classification performance to each set of models to be evaluated.

In this case the first classification model uses a neural network algorithm to produce a confusion matrix in Figure 4.

accuracy: 95.00%

	true BERHI	true DEGL...	true DOKOL	true IRAQI	true ROTA...	true SAFAVI	true SOGAY	class preci...
pred. BERHI	16	0	0	0	0	0	0	100.00%
pred. DEG...	0	9	0	0	0	0	1	90.00%
pred. DOK...	0	0	14	0	0	1	0	93.33%
pred. IRAQI	0	0	0	16	0	0	0	100.00%
pred. ROT...	0	0	0	0	15	0	1	93.75%
pred. SAFAVI	0	0	0	0	0	16	0	100.00%
pred. SOGAY	0	1	1	0	0	0	9	81.82%
class recall	100.00%	90.00%	93.33%	100.00%	100.00%	94.12%	81.82%	

Fig. 4. Neural network model confusion matrix

Based on the resulting confusion matrix in Figure 4, the first model produces an accuracy value of 95%. The second is an evaluation of the classification model using a neural network algorithm plus bagging optimization. Following are the results of the confusion matrix from the second model.

accuracy: 96.00%

	true BERHI	true DEGL...	true DOKOL	true IRAQI	true ROTA...	true SAFAVI	true SOGAY	class preci...
pred. BERHI	16	0	0	0	0	0	0	100.00%
pred. DEG...	0	9	0	0	0	0	1	90.00%
pred. DOK...	0	0	14	0	0	0	0	100.00%
pred. IRAQI	0	0	0	16	0	0	0	100.00%
pred. ROT...	0	0	0	0	15	0	1	93.75%
pred. SAFAVI	0	0	0	0	0	17	0	100.00%
pred. SOGAY	0	1	1	0	0	0	9	81.82%
class recall	100.00%	90.00%	93.33%	100.00%	100.00%	100.00%	81.82%	

Fig. 5. Neural network optimization model confusion matrix using bagging

Based on the resulting confusion matrix in Figure 5, the second model produces an accuracy value of 96%. The results of the evaluation of the classification process using the Neural Network algorithm and the use of the bagging technique in the Neural Network algorithm produce the following accuracy comparisons.

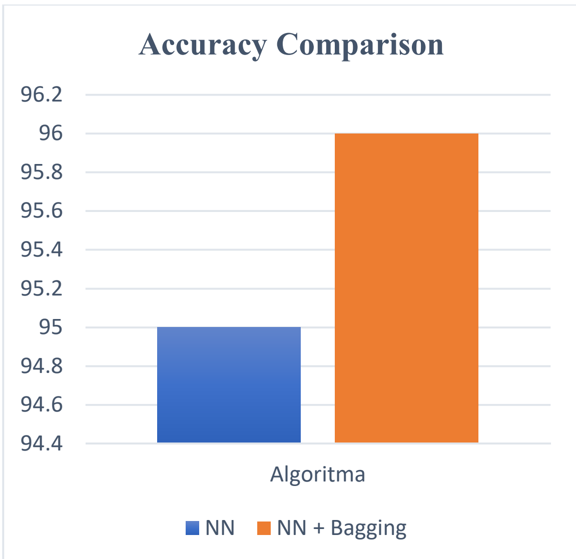

Fig. 6. Comparison of accuracy of nn and nn + bagging algorithms

Based on the comparison of the accuracy values in Figure 6, it can be concluded that the accuracy value of the neural network algorithm model is smaller than the optimization model using the bagging algorithm. So it can be said that the bagging algorithm can improve the accuracy of neural network classification by 1%.

IV. CONCLUSIONS

Based on the results of the research that has been done, it can be concluded that the neural network algorithm can be optimized using one technique, namely bagging. This can be proven by the results of a comparison of the two using public datasets to find the level of accuracy of both. With a dataset of 898 and taken 100 data as data testing. Neural networks optimized using bagging techniques can achieve an accuracy rate of 96% while using only a neural network produces an accuracy of 95%. So there is an accuracy gap of 1%. Suggestions for further research are that optimization can be carried out using other techniques or can be tested with other datasets.

REFERENCES

[1] A. H. Yunial, "Optimization Analysis of Support Vector Machine Classification Algorithms, Decision Trees, and Neural Networks Using Adaboost and Bagging," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 3, p. 247, Sep. 2020, doi: 10.32493/informatika.v5i3.6609.

[2] S. B. Xu, S. Y. Huang, Z. G. Yuan, X. H. Deng, and K. Jiang, "Prediction of the Dst Index with Bagging Ensemble-learning Algorithm," *Astrophys J Suppl Ser*, vol. 248, no. 1, p. 14, May 2020, doi: 10.3847/1538-4365/ab880e.

[3] O. Vigo, S. Lukas, D. Krisnadi, P. Widjaja, and P. Korespondensi, "Comparison Of Bagging And Adaboost Performance For Multi-Class Data Classification," vol. 7, no. 2, 2022.

[4] U. Michelucci and F. Venturini, "Estimating Neural Network's Performance with Bootstrap: A Tutorial," *Mach Learn Knowl Extr*, vol. 3, no. 2, pp. 357–373, Mar. 2021, doi: 10.3390/make3020018.

[5] C. O. Siti Fatihah *et al.*, "The effect of aggregating bootstrap on the accuracy of neural network system for islamic investment prediction," *Universal Journal of Accounting and Finance*, vol. 9, no. 4, pp. 604–612, Aug. 2021, doi: 10.13189/ujaf.2021.090408.

[6] R. du Plooy and P. J. Venter, "A Comparison of Artificial Neural Networks and Bootstrap Aggregating Ensembles in a Modern Financial Derivative Pricing Framework," *Journal of Risk and Financial Management*, vol. 14, no. 6, p. 254, Jun. 2021, doi: 10.3390/jrfm14060254.

[7] M. Smaida, "Bagging of Convolutional Neural Networks for Diagnostic of Eye Diseases," 2020.

[8] G. Arther Sandag, "A Prediction Model Of Company Health Using Bagging Classifier," vol. 6, no. 1, 2020, doi: 10.33480/jitk.v6i1.1390.

[9] T. Ginanjar Laksana, J. Informatika, and S. di Tinggi Telematika Telkom Purwokerto Jl Panjaitan, "Comparison Of C4.5 Algorithm With C4.5 Bagging In Predicting And Analysis Of Data Selector Set Of Engine Lubricant Oil (Oil) On Two-Wheel Vehicles," 2018.

[10] A. Kurniawan and A. Prihandono, "Application Of Bagging Techniques To Improve Classification Accuracy In The Naive Bayes Algorithm In Determining Professional Bloggers," 2020.

[11] H. J. Yoon *et al.*, "Accelerated training of bootstrap aggregation-based deep information extraction systems from cancer pathology reports," *J Biomed Inform*, vol. 110, Oct. 2020, doi: 10.1016/j.jbi.2020.103564.

[12] Y. L. Nursimpati and A. Saifudin, "Application of Naïve Bayes-Based Bagging Technique for Student Admissions Selection," vol. 4, no. 2, pp. 2622–4615, 2019, [Online]. Available: http://openjournal.unpam.ac.id/index.php/informatika65

[13] S. Al-Dahidi, P. Baraldi, E. Zio, and M. Lorenzo, "Bootstrapped ensemble of artificial neural networks technique for quantifying uncertainty in prediction of wind energy production," *Sustainability (Switzerland)*, vol. 13, no. 11, Jun. 2021, doi: 10.3390/su13116417.

[14] S. Mutmainnah, G. Abdurrahman, H. Azizah, and A. Faruq, "C4.5 Algorithm Optimization Using Bagging Techniques On Gold Karat Level Data," 2020. Accessed: Nov. 29, 2022. [Online]. Available: http://repository.unmuhjember.ac.id/4416/

[15] J. Abdollahi, B. Nouri-Moghaddam, and M. Ghazanfari, "Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases)."

[16] M. O. Elish and K. Elish, "An Empirical Comparison of Resampling Ensemble Methods of Deep Learning Neural Networks for Cross-Project Software Defect Prediction," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 3, pp. 201–209, Jun. 2021, doi: 10.22266/ijies2021.0630.18.

[17] I. Possebon, A. Silva, L. Granville, A. Schaeffer-Filho, and A. Marnerides, "Improved Network Traffic Classification Using Ensemble Learning," *IEEE Symposium on Computers and Communications (ISCC)*, 2019.

[18] A. Dželihodžić, D. Donko, and J. Kevrić, "Improved credit scoring model based on bagging neural network," *Int J Inf Technol Decis Mak*, vol. 17, no. 6, pp. 1725–1741, Nov. 2018, doi: 10.1142/S0219622018500293.

[19] D. Triana and S. Osowski, "Bagging and boosting techniques in prediction of particulate matters," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 68, no. 4, pp. 1207–1215, Oct. 2020, doi: 10.24425/bpasts.2020.134659.

[20] R. Singh, "Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 3970–3976, Jun. 2020, doi: 10.30534/ijatcse/2020/221932020.