Jurnal Teknologi Informasi dan Pendidikan
Volume 18, No. 2, September 2025
https://doi.org/10.24036/jtip.v18i2.981

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Application of Bagging Ensemble Learning on Naïve Bayes Algorithm to Predict Coronary Heart Disease

I Gusti Agung Satria Nugraha[1]*✉, I Gede Aris Gunadi[1], Luh Joni Erawati Dewi[1]
[1]Department of Computer Science, Universitas Pendidikan Ganesha, Bali, Indonesia
Corresponding Author: agung.satria@undiksha.ac.id

Article Information

Article history:

No. 981
Rec. July 01, 2025
Rev. August 19, 2025
Acc. August 19, 2025
Pub. August 19, 2025
Page. 943 – 954

Keywords:

- *Coronary Heart Disease*
- *Data Mining*
- *Naïve Bayes*
- *Bagging*
- *Confusion Matrix*

ABSTRACT

Cardiovascular health is vital, with heart disease, particularly Coronary Heart Disease (CHD), being a significant health concern in Indonesia. The 2023 Indonesian Health Survey reported 877,531 cases of heart disease. Traditional CHD diagnosis is often costly and invasive. Therefore, machine learning-based classification has emerged as a promising alternative for enhancing the accuracy and efficiency of detection. This study aims to predict CHD using a hybrid approach combining the Naïve Bayes algorithm with the Bagging ensemble method. Naïve Bayes was selected for its computational efficiency and effectiveness with high-dimensional data, while Bagging was employed to mitigate its inherent weaknesses by reducing variance and increasing prediction stability. The CRISP-DM methodology was applied to a secondary dataset of 462 rows from Kaggle. The research process included data preprocessing, method implementation, and evaluation using a confusion matrix. Results show the Bagging method with n=2 estimators achieved optimal performance, with 76.34% accuracy, 65.00% precision, and an f1-score of 70.27%. This study demonstrates that ensemble techniques can effectively improve the accuracy and stability of CHD prediction models, offering a reliable and low-cost solution for initial screening.

How to Cite:

Nugraha, I. G. A. S., & et al. (2025). Application of Bagging Ensemble Learning on Naïve Bayes Algorithm to Predict Coronary Heart Disease. Jurnal Teknologi Informasi Dan Pendidikan, 18(2), 943-954. https://doi.org/10.24036/jtip.v18i2.981

1. INTRODUCTION

Cardiovascular health is paramount to human quality of life, ensuring the vital distribution of oxygen and essential nutrients throughout the body's cells. Disruptions within the cardiovascular system directly impact an individual's well-being and life expectancy. Recent data from the 2023 Indonesian Health Survey [1] reveals a significant prevalence of heart disease, with 877,531 diagnosed cases and an average of 0.85% of the population. Among various cardiac disorders, Coronary Heart Disease (CHD) stands out as a major concern. CHD, characterized by the narrowing and blockage of coronary blood vessels due to the accumulation of plaque and pollutants, severely impedes energy transport and oxygen supply, thereby threatening bodily homeostasis [2].

Traditional CHD diagnosis often involves invasive and high-cost procedures, presenting a substantial challenge to healthcare systems [3]. This is in line with research [4], machine learning has emerged as a promising solution to improve the accuracy and efficiency of medical detection. Through data mining, valuable insights can be extracted from large medical datasets, enabling the identification of patterns and relationships that can support more precise clinical decision-making [5]. Specifically, classification methods within data mining are highly relevant for assigning class labels to data samples, such as identifying individuals at risk of CHD, thereby bolstering the reliability of the generated results.

The Naïve Bayes algorithm is a widely utilized probability-based classification method, recognized for its computational efficiency and high accuracy, particularly with large datasets [6]. This algorithm calculates probabilities based on the frequency and combination of values within a dataset, making the naive assumption that each feature is independent of other features if the outcome is known [7]. However, this algorithm has limitations, including its aforementioned assumption of feature independence and suboptimal handling of continuous attributes [8] [9]. To overcome the limitations of single algorithms and improve overall model performance, ensemble learning has proven highly effective. Bagging (Bootstrap Aggregating), as a key ensemble technique, operates by creating multiple versions of the training dataset through sampling with replacement to train several individual classifiers [10]. This approach effectively reduces variance and enhances prediction stability [11].

Previous research has demonstrated the significant potential of ensemble learning in the medical domain. For instance, applying Bagging and AdaBoost to the C4.5 algorithm for stroke prediction yielded notable improvements. While the standalone C4.5 algorithm achieved an accuracy of 92.87%, the use of Bagging boosted accuracy to 95.02%, and AdaBoost reached 94.63%, indicating substantial increases of 3% and 2%, respectively [12]. In another study, a Bagging method combining K-Nearest Neighbor, Support Vector Machine, and Naïve Bayes for predicting stunting in toddlers achieved an accuracy of 89.77% [13]. Given Bagging's proven effectiveness in enhancing classification performance, and considering the unique characteristics of the Naïve Bayes algorithm, this study aims to evaluate how the combination of the Bagging method with the Naïve Bayes algorithm can be optimized for Coronary Heart Disease prediction. Therefore, this research will comprehensively investigate the implementation of Bagging Ensemble Learning on the

Naïve Bayes Algorithm for Predicting Coronary Heart Disease, focusing on improving key classification metrics such as accuracy (the proportion of correct predictions from the total data), precision (the model's ability to not misclassify positive results), and f1-score (the harmonic mean of precision and recall which provides a balanced picture of model performance).

2. RESEARCH METHOD

This study employed the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, a widely recognized reference model that systematically outlines the data mining project lifecycle from planning to implementation [14][15]. This method comprises structured phases that facilitate data management and is frequently utilized in the field of data science due to its clear, efficient workflow, and its ability to reduce research costs [16] [17]. The method can be seen in Figure 1.

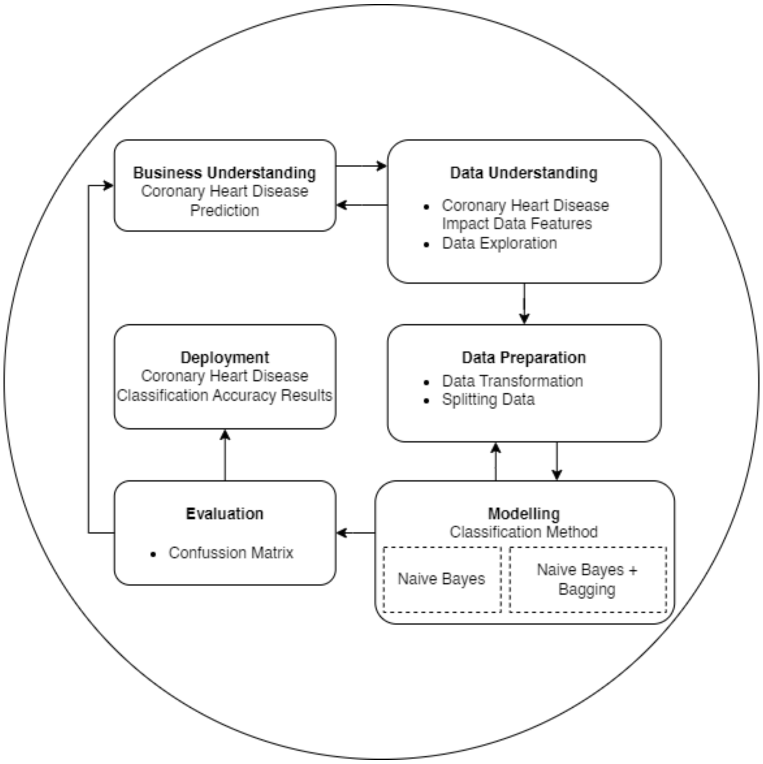


Figure 1. CRISP-DM methodology

2.1 Bussiness Understanding

In the Business Understanding phase, the primary objective of this research is to develop a predictive model for detecting coronary heart disease (CHD) that can assist healthcare institutions in clinical decision-making. The primary objectives are to filter relevant data related to CHD risk and to identify significant patterns for prediction [18]. The development and diagnosis of CHD are influenced by both internal and external factors. Internal factors include genetic conditions and patient medical history, such as high blood pressure, diabetes, obesity, and lifestyle habits like smoking and physical activity levels. External factors encompass environmental conditions, including air pollution, social stress, education levels, access to healthcare services, and dietary patterns related to the patient's lifestyle. In line with previous research [19], this study investigates the relationships between health factors from patient medical records and CHD using Naïve Bayes algorithm and a combination of Naïve Bayes with Bagging. This approach is designed to support medical decision-making, assist in preventative measures by identifying high-risk individuals for early intervention, and support ongoing management by monitoring the effectiveness of therapy and personalizing treatment plans.

2.2 Data Understanding

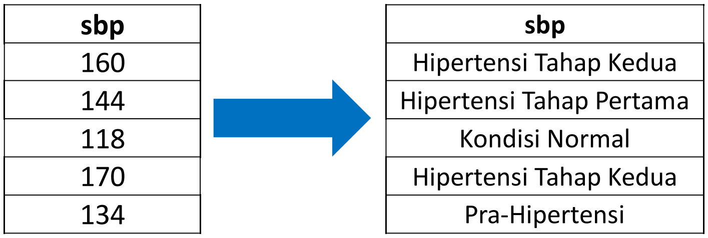
In this phase, we initiated our data understanding by gathering the initial dataset and reviewing its structure to identify potential issues within the Coronary Heart Disease (CHD) dataset. The data was sourced from Kaggle, specifically the "Coronary Heart Disease" dataset, comprising 462 rows and 9 attributes. The features identified as influential for CHD prediction include *sbp, tobacco, ldl, adiposity, famhist, typea, obesity, alcohol, and age*. The insights gained from this exploration were crucial in determining the criteria for our model, which involved grouping data based on the values of each feature within the dataset [20]. A detailed description of the features found in the dataset is provided in Table 1.

Table 1. Description features dataset

Features	Description	Type
sbp	Systolic blood pressure	Numeric
tobacco	Yearly tobacco use (in kilograms)	Numeric
ldl	Low-density lipoprotein (LDL) cholesterol level	Numeric
adiposity	Level of body fat	Numeric
famhist	Family history of CHD (0 = No, 1 = Yes)	Categorical
typea	Type A behavior/personality score	Numeric
obesity	Body Mass Index (BMI)	Numeric
alcohol	Alcohol consumption	Numeric
age	Age of the patient	Numeric
chd	Coronary Heart Disease diagnosis (0 = No, 1 = Yes)	Categorical

2.3 Data Preparation

The Coronary Heart Disease dataset, obtained from Kaggle in CSV format, served as the foundation for this study. This dataset comprises 462 patient records, each with 9 attributes, including 1 target attribute. For subsequent analysis and model development, the dataset was processed using Python. Prior to classification, a series of data preprocessing steps were performed. It's important to note that while some researchers consider data mining synonymous with knowledge discovery in databases, data mining is actually just one crucial step in that broader process [21]. For instance, quantitative attributes like '*sbp*' (systolic blood pressure) were converted into categorical ranges. The specific conversion criteria for each attribute are detailed in Figure 2. This transformation was deemed necessary to align with the chosen classification approach. For this study, 80% of the dataset was used for training and 20% for testing purposes.



sbp
160
144
118
170
134

sbp
Hipertensi Tahap Kedua
Hipertensi Tahap Pertama
Kondisi Normal
Hipertensi Tahap Kedua
Pra-Hipertensi

Figure 2. Data Transformation

2.4 Modelling

In this research phase, the dataset was analyzed using two distinct models: Naïve Bayes and a combined Naïve Bayes with Bagging approach. We conducted a comprehensive comparison between the actual testing data and the predicted outcomes generated by both the trained Naïve Bayes model and the integrated Naïve Bayes with Bagging model. This comparison aimed to assess their respective performances in predicting the target variable on the testing dataset.

2.5 Evaluation

In this phase, the performance of the developed models was rigorously evaluated to ensure they met the predefined research objectives. The classification report provided a comprehensive summary of the predicted data's evaluation, offering insights into accuracy, precision, recall, and F1-score for each class [22]. This report, along with the confusion matrix, provides a detailed breakdown of prediction outcomes. Specifically, the confusion matrix categorizes predictions into four unique combinations of predicted results and actual conditions: true positive, true negative, false positive, and false negative [23]. The confusion matrix served as a crucial performance evaluation tool, presenting a detailed overview of the models' prediction outcomes and was instrumental in analyzing how accurately each model recognized and correctly classified every category within the dataset. The final step involved a comparative analysis between the Naïve Bayes model and the combined Naïve Bayes with Bagging model to determine which approach yielded superior performance in predicting coronary heart disease.

2.6 Deployment

This research demonstrates that the combination of the Naïve Bayes algorithm with Bagging is a superior method for predicting the likelihood of an individual suffering from coronary heart disease. This research [24] supported by a specific review of how machine learning predictive models are an effective and promising solution for diagnosing chronic diseases, including heart disease. Specifically, the Bagging method, with an estimator configuration of n=2, achieved an optimal accuracy of 76.34%. This performance indicates that Bagging effectively reduces model variance and enhances its generalization capability for coronary heart disease prediction. For instance, by leveraging the trained Bagging model, if patient data indicates a low-risk profile, it can be predicted that the individual is likely not suffering from coronary heart disease.

3. RESULTS AND DISCUSSION

This study explores the capabilities of two machine learning approaches, namely the Naïve Bayes algorithm and a combination of Naïve Bayes with Bagging, in predicting the occurrence of coronary heart disease. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics on a dataset comprising 462 instances.

3.1 Naive Bayes Algorithm Classification

The Naïve Bayes classification model was tested on the pre-processed dataset, comprising 462 instances. This evaluation aimed to assess the model's performance in accurately classifying the target variable.

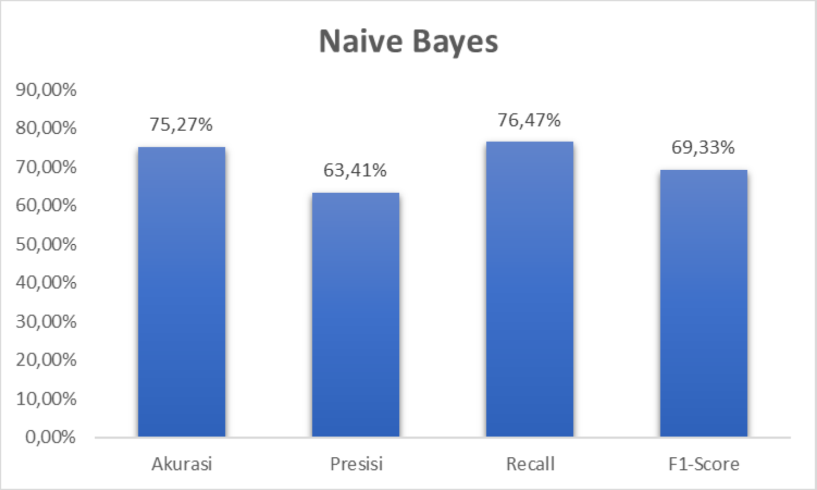


Figure 3. Naïve Bayes Algorithm Classification

As illustrated in Figure 3, the Naïve Bayes algorithm yielded an accuracy of 75.27%, indicating a reasonably good prediction capability. Additionally, the model achieved a precision of 63.41%, a recall of 76.47%, and an F1-score of 69.33%. The higher recall value compared to precision suggests that the model is quite effective at correctly identifying positive class instances, despite some remaining misclassifications. Overall, these results indicate that the Naïve Bayes method provides adequate classification performance on the dataset used in this study.

3.2 Combined classification of the Naive Bayes Algorithm with bagging

The classification model was also tested using a combination of the Naïve Bayes algorithm and Bagging on the pre-processed dataset of 462 instances. The Bagging algorithm was implemented by first generating bootstrap samples using the sampling with replacement method, creating new training data subsets, each with the same number of instances as the original dataset. In this study, 10 such subsets were generated. Statistically, approximately 63.2% of unique data points appeared in each sample, with the remainder being duplicates or unselected instances. Subsequently, in the modeling phase, each training data subset was used to train an individual Naïve Bayes classification model. Finally, during the aggregation/voting prediction stage, the predictions from all 10 individual Naïve Bayes models were combined using a majority voting method to determine the final class output.

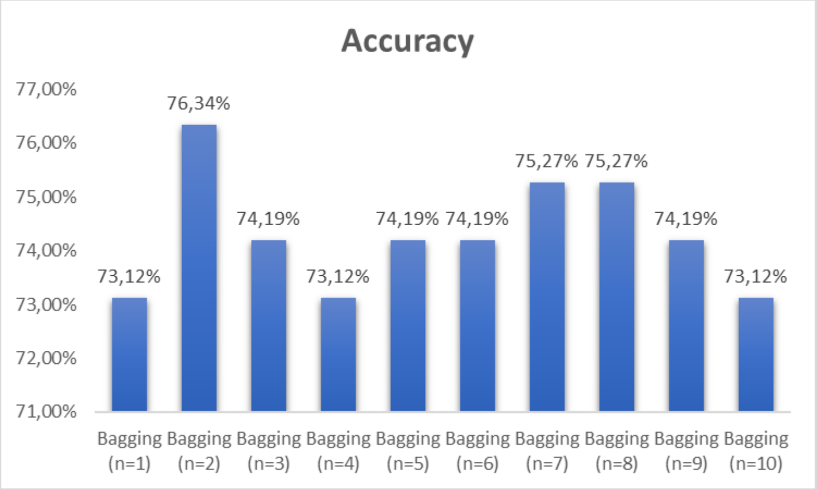


Figure 4. Accuracy of Naive Bayes and Bagging Combination

As shown in Figure 4, the classification accuracy using the combined Naïve Bayes and Bagging algorithm was evaluated across a range of estimator numbers (n) from 1 to 10. The results consistently demonstrated that Bagging improved accuracy compared to using a single base model (n=1). The highest accuracy of 76.34% was achieved with n=2 estimators. After peaking at n=2, the accuracy fluctuated, generally remaining within the 73% to 75%

range. However, adding too many estimators (e.g., at n=9 or n=10) led to a decrease in accuracy. This suggests an optimal point exists, and excessive estimators could potentially lead to minor overfitting or limitations in addressing bias.

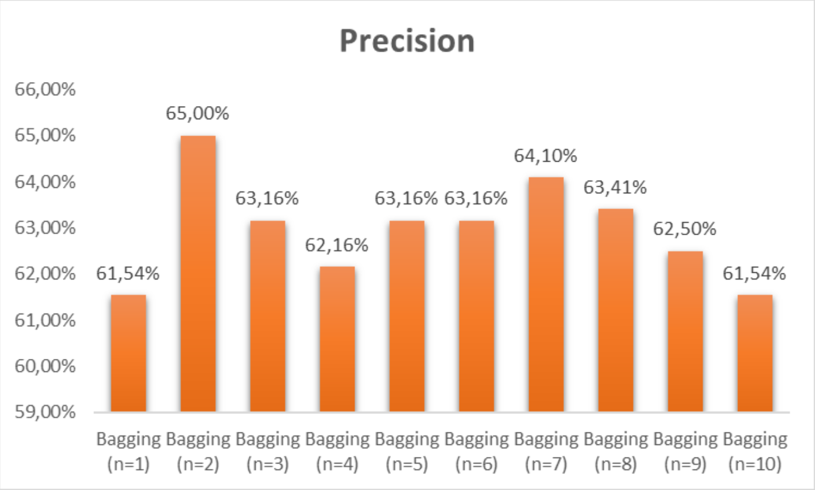


Figure 5. Precision of Naive Bayes and Bagging Combination

As depicted in Figure 5, the precision results for the combined Naïve Bayes and Bagging classification show the most significant improvement at the Bagging configuration with n=2 estimators, achieving a peak precision of 65.00%. This represents an increase of approximately 3.46% compared to n=1. While numerically modest, this improvement is considered effective in reducing false positives and enhancing the accuracy of positive class identification compared to other estimator configurations. This finding underscores the existence of an optimal point for precision when using Bagging, suggesting that a higher number of estimators does not consistently translate to better performance.

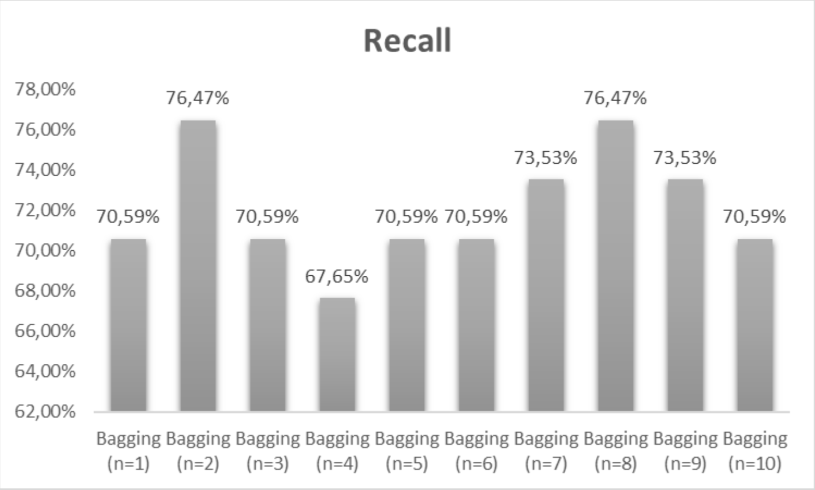


Figure 6. Recall of Naive Bayes and Bagging Combination

As shown in Figure 6, the recall results for the combined Naïve Bayes and Bagging classification indicate that Bagging is effective in reducing false negatives, meaning the model is better at identifying all true positive instances within the dataset. Two identical optimal recall values of 76.47% were observed at n=2 and n=8 estimators. This phenomenon likely stems from the variations in bootstrap samples and how these subsets influence the base model's ability to capture difficult-to-recognize positive cases.

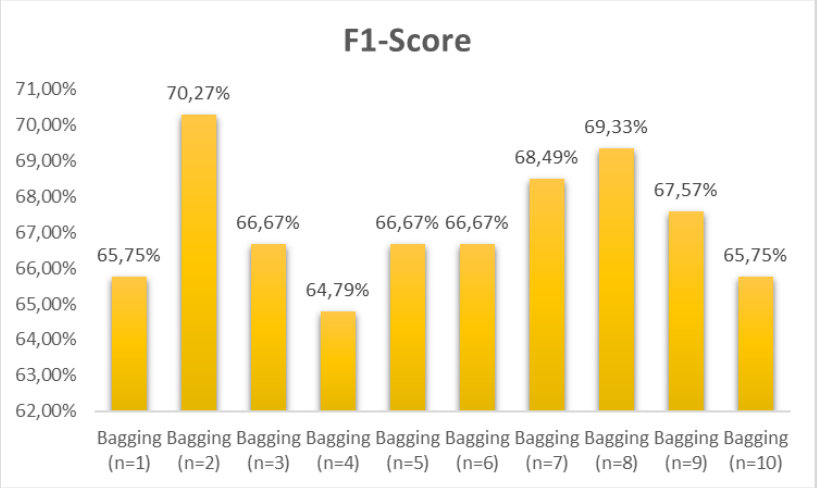


Figure 7. F1-score of Naive Bayes and Bagging Combination

As presented in Figure 4.5, the F1-score for the combined Naïve Bayes and Bagging classification shows that the Bagging model with n=2 estimators achieved a score of 70.27%. This indicates that the addition of just one estimator can significantly impact achieving a good balance between precision (reducing false positives) and recall (reducing false negatives). Similar to the patterns observed in previous metrics (accuracy and precision), the F1-score tended to decrease with a higher number of estimators.

3.3 Comparison Test Results

After evaluating individual classification metrics, a comprehensive comparison of all obtained results was performed, as presented in Table 2.

Table 2. Comparison test results

Method	Accuracy	Precision	Recall	F1-Score
Naive Bayes	75,27%	63,41%	76,47%	69,33%
Bagging (n=1)	73,12%	61,54%	70,59%	65,75%
Bagging (n=2)	76,34%	65,00%	76,47%	70,27%
Bagging (n=3)	74,19%	63,16%	70,59%	66,67%
Bagging (n=4)	73,12%	62,16%	67,65%	64,79%
Bagging (n=5)	74,19%	63,16%	70,59%	66,67%
Bagging (n=6)	74,19%	63,16%	70,59%	66,67%
Bagging (n=7)	75,27%	64,10%	73,53%	68,49%
Bagging (n=8)	75,27%	63,41%	76,47%	69,33%

Bagging (n=9)	74,19%	62,50%	73,53%	67,57%
Bagging (n=10)	73,12%	61,54%	70,59%	65,75%

Overall, the implementation of Bagging with Naïve Bayes demonstrated varied performance improvements depending on the number of estimators (n). Bagging with n=2 emerged as the most optimal configuration across all evaluated metrics, achieving an accuracy of 76.34%, precision of 65.00%, recall of 76.47%, and an F1-score of 70.27%.

Analyzing the classification results with optimal values, the accuracy metric clearly shows Bagging at its peak, effectively reducing variance and enhancing the model's generalization capability. For precision, Bagging proved highly effective in mitigating false positives and improving the relevance of positive predictions. However, a notable observation was the identical highest recall values (76.47%) found for both the standalone Naïve Bayes algorithm and the combined Naïve Bayes with Bagging at n=2 and n=8 estimators. This indicates that while Bagging successfully maintained the true positive detection ability of Naïve Bayes, even after potential initial decreases due to bootstrap samples, it did not consistently improve recall beyond what was already achieved by the base Naïve Bayes algorithm. Regarding the F1-score, Bagging, at its optimal configuration (n=2), achieved a slightly better balance between precision and recall. Nevertheless, it's important to note that most other Bagging configurations yielded lower F1-scores compared to the standalone Naïve Bayes algorithm.

4. CONCLUSION

This research successfully applied the Bagging ensemble learning method to Naïve Bayes classification for predicting coronary heart disease. The dataset utilized was sourced from Kaggle, titled "Coronary Heart Disease," comprising 462 data instances and including key features such as sbp, tobacco, ldl, adiposity, famhist, typea, obesity, alcohol, and age. The Bagging implementation involved independently training multiple base Naïve Bayes classification models on different bootstrap samples of the original dataset, followed by combining their predictions through a majority voting process.

The performance comparison revealed that Bagging with 2 estimators (n=2) achieved the most optimal results, yielding an accuracy of 76.34%, precision of 65.00%, recall of 76.47%, and an F1-score of 70.27%. In contrast, the standalone Naïve Bayes algorithm achieved an accuracy of 75.27%, precision of 63.41%, recall of 76.47%, and an F1-score of 69.33%. Thus, the application of the Bagging method to the Naïve Bayes algorithm effectively improved classification performance, particularly across accuracy, precision, and F1-score metrics. This process proved successful in reducing the variance of the Naïve Bayes algorithm and making the predictions more robust. Although Naïve Bayes relies on a naive assumption of feature independence, Bagging can help mitigate this limitation by providing diverse data perspectives to each base model, leading to better collective decisions and addressing the weaknesses of a single-assumption model.

REFERENCES

[1] B. K. P. Kesehatan, "Survei Kesehatan Indonesia (SKI) 2023," 2023.
[2] E. Erdania, M. Faizal, and R. B. Anggraini, "Faktor – Faktor Yang Berhubungan Dengan Kejadian Penyakit Jantung Koroner (PJK) Di RSUD Dr. (H.C.) Ir. SOEKARNO PROVINSI BANGKA BELITUNG TAHUN 2022," *J. Keperawatan*, vol. 12, no. 1, pp. 17–25, Jun. 2023, doi: 10.47560/kep.v12i1.472.
[3] R. T. Wahyuni, E. Witcahyo, and Y. T. Herawati, "Hubungan Karakteristik Pasien, Prosedur, dan Penyakit Penyerta Dengan Biaya Langsung Medis Pada Pasien Rawat Inap Jantung Koroner," *J. Ekon. Kesehat. Indones.*, vol. 8, no. 1, p. 1, Jul. 2023, doi: 10.7454/eki.v8i1.6240.
[4] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, doi: 10.3390/healthcare10030541.
[5] S. Yang, "A Study of Heart Disease Diagnosis Using Machine Learning and Data Mining," *J. Clin. Med. Res.*, vol. 5, no. 4, p. 565, Dec. 2024, doi: 10.32629/jcmr.v5i4.3135.
[6] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes," *MDP Student Conf.*, vol. 2, no. 1, pp. 297–305, Apr. 2023, doi: 10.35957/mdp-sc.v2i1.4388.
[7] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovensional dalam Prediksi Komplikasi Hipertensi," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 6, no. 1, pp. 37–49, Apr. 2024, doi: 10.35746/jtim.v6i1.488.
[8] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, Jun. 2021, doi: 10.29207/resti.v5i3.3067.
[9] H. Mukaromah and Wasilah, "Komparasi Teknik Bagging Dan Adaboost Pada Decision Tree Dan Naive Bayes Untuk Prediksi Stroke," *J. Penelit. Ilmu dan Teknol. Komput.*, vol. 16, no. 1, pp. 167–180, 2024.
[10] F. Kamalov, A. Elnagar, and H. H. Leung, "Ensemble Learning with Resampling for Imbalanced Data," 2021, pp. 564–578. doi: 10.1007/978-3-030-84529-2_48.
[11] A. Ridwan, "Penerapan Teknik Bagging Pada Algoritma Naive Bayes Dan Algoritma C4.5 Untuk Mengatasi Ketidakseimbangan Kelas," *J. Bisnis Digit. dan Sist. Inf.*, vol. 1, no. 1, pp. 63–70, 2020.
[12] N. D. Saputri, K. Khalid, and D. Rolliawati, "Komparasi Penerapan Metode Bagging dan Adaboost pada Algoritma C4.5 untuk Prediksi Penyakit Stroke," *J. Sist. Inf.*, vol. 11, no. 3, pp. 567–577, 2022, doi: https://doi.org/10.32520/stmsi.v11i3.
[13] G. N. Masacgi and M. S. Rohman, "Optimasi Model Algoritma Klasifikasi menggunakan Metode Bagging pada Stunting Balita," *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 455–464, Dec. 2023, doi: 10.29408/edumatic.v7i2.23812.
[14] M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11,

no. 1, p. 1, Apr. 2022, doi: 10.23887/janapati.v11i1.45282.

[15] I. G. Hendrayana, D. G. H. Divayana, and M. W. A. Kesiman, “Komparasi Metode Svm, K-Nn Dan Nbc Pada Analisis Sentimen,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 1, pp. 191–198, Jan. 2023, doi: 10.35870/jimik.v4i1.157.

[16] A. Rianti, N. W. A. Majid, and A. Fauzi, “CRISP-DM: Metodologi Proyek Data Science,” in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, 2023, pp. 107–114.

[17] D. B. Saputra, V. Atina, and F. E. Nastiti, “Penerapan Model Crisp-Dm Pada Prediksi Nasabah Kredit Menggunakan Algoritma Random Forest,” *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 240–247, Jul. 2024, doi: 10.36080/idealis.v7i2.3244.

[18] D. Larassati, A. Zaidiah, and S. Afrizal, “Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 2, pp. 533–546, May 2022, doi: 10.29100/jipi.v7i2.2842.

[19] P. A. Sihotang and D. Sitanggang, “Penerapan Metode Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Jantung,” *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 899, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1535.

[20] A. E. Karrar and R. Elarabi, “An Intelligent Machine Learning-Based System for Predicting Heart Disease Using Mixed Feature Creation Technique,” in *Journal of Physics: Conference Series*, 2023, pp. 353–367. doi: 10.1007/978-3-031-36258-3_31.

[21] Z. Khoshgoftar, M. Babaee, A. K. Rouzbahani, and M. Kalantarion, “Educational data mining in medical education: A five-level approach,” *J. Educ. Health Promot.*, vol. 14, no. 1, Jan. 2025, doi: 10.4103/jehp.jehp_1339_23.

[22] O. B. Akanbi, “Prediction of Heart Disease Risk among Patients in Federal Medical Centre, Abeokuta Using Naïve Bayes,” *Asian J. Probab. Stat.*, vol. 26, no. 10, pp. 46–63, Sep. 2024, doi: 10.9734/ajpas/2024/v26i10658.

[23] K. S. Y. Pande, D. G. H. Divayana, and G. Indrawan, “Comparative analysis of naïve bayes and knn on prediction of forex price movements for gbp/usd currency at time frame daily,” in *The 3rd International Conference on Vocational Education and Technology (IConVET)*, 2021.

[24] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. Amenta, “Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis,” *J. Pers. Med.*, vol. 10, no. 2, p. 21, Mar. 2020, doi: 10.3390/jpm10020021.