ILKOM Jurnal Ilmiah Vol. 15, No. 2, August 2023, pp. 344-352
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

A Soft Voting Ensemble Classifier to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients

Arif Munandar [a,1]; Wiga Maulana Baihaqi [a,2,*]; Ade Nurhopipah [a,3]
[a] *Universitas Amikom Purwokerto, Jl. Letjend Pol. Soemarto No.127, Banyumas, 53127, Indonesia*
[1] *arifmunandarr404@gmail.com;* [2] *wiga@amikompurwokerto.ac.id;* [3]*ade_nurhopipah@amikompurwokerto.ac.id*
* *Corresponding author*

Article history: Received February 08, 2023; Revised June 16, 2023; Accepted July 28, 2023; Available online August 16, 2023

Abstract

Cardiovascular disease is one of the deadliest diseases, claiming around 17 million lives worldwide each year. According to data from the WHO, more than four out of five deaths from cardiovascular disease are caused by heart attacks and strokes, and one-third of these deaths occur prematurely in people under the age of 70. Machine learning approaches can be used to detect the disease. This research aims to improve the prediction model of cardiovascular heart failure patient survival using C4.5, KNN, Logistic Regression algorithms, and the ensemble learning method of voting classifier. Based on the testing results, each model showed a significant increase in accuracy in the 70:30 ratio. Logistic regression and C4.5 achieved the same accuracy, 89.47%, KNN obtained 91.23%, and voting classifier experienced a considerable improvement, reaching 94.74%. In testing with ratios of 90:10, 80:20, and 70:30, KNN demonstrated high accuracy but had significant overfitting, with a difference of 7-9% between training and testing accuracy scores in the 90:10 and 80:20 ratios. On the other hand, voting classifier showed stable performance in the 70:30 ratio, with an accuracy difference between training and testing scores below 1%. The conclusion of this research is that the voting classifier can assist the performance improvement of algorithms for classifying the survival expectancy of cardiovascular heart failure patients into 'Survived' or 'Deceased', compared to logistic regression, KNN, and C4.5.

Keywords: Cardiovascular; C4.5; Ensemble Learning; K-Nearest Neighbors; Logistic Regression; Machine Learning; Voting Classifier.

Introduction

Cardiovascular disease is one of the deadliest diseases, taking about 17 million people's lives worldwide every year [1-4]. Cardiovascular disease is a disease attacking the heart and blood vessels, including coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other heart diseases. According to WHO (World Health Organization) data, more than four out of five deaths from cardiovascular disease are caused by heart attacks, and strokes, and one-third of those deaths occur prematurely in people under 70 years of age [2][4].

According to estimates, 17.9 million people worldwide died in 2019 from cardiovascular disease, accounting for 32% of all deaths. Then 85% of those deaths are caused by strokes and heart attacks. Therefore, prevention such as the detection of heart disease in children must be intensified at all times so that the patient's condition can be improved with treatment and counseling [4].

Based on the presentation of the facts above and the explanation of the problem, research in detecting cardiovascular heart disease is needed. Predicting a patient's life expectancy after being discharged from the hospital due to heart failure is a major challenge [5]. Machine learning approaches can be used to detect disease [6]. In the heart failure dataset, it is classified as an imbalance class, where there are more positive data (alive) than negative data (death) [7]. Class imbalance occurs because the number of majority classes far exceeds the number of minority classes. This will affect the accuracy of the results [6][8].

http://dx.doi.org/10.33096/ilkom.v15i2.1632.344-353

Several studies have been conducted regarding the survival prediction in patients with cardiovascular heart failure. Research conducted by [9] on the prediction of heart failure using the Support Vector Machine (SVM) obtained an accuracy of 90.11% with a linear kernel, then in the study [3] using the KNN algorithm obtained an accuracy value of 94.92% with a value of k=7, then in the study [10] using the random forest, SVM, gradient boosting, XGBoost, and LightGBM algorithm models obtained the best accuracy value which was 80% on XGBoost and SVM, then in the study [11] using hyper-parameter tuning on XGBoost obtained its best AUC value of 94%, then in the study [12] using the best first feature selection on the Random Forest algorithm produced the best accuracy of 83%, in the study [2] using random forest got the best accuracy value of 85% with optimization techniques using k-fold and GridSearchCV, in the study [13] comparing the SVM algorithm, KNN, and random forest got the best accuracy on svm and random forest at 96%, and the last one in the study [14] compared the decision tree, naïve bayes, and random forest algorithms to get 0.70% decision tree accuracy values, 0.72% naïve bayes, and 0.75% random forest, respectively.

Some studies have been conducted to improve accuracy performance. The study [6] overcome imbalance classes using the SMOTE method in improving stroke detection performance with 91% accuracy results, then the study [7] used the Weighted Random Forest method in improving the performance of the classification of predictions of survival of heart failure patients with an accuracy result of 90.3%, then in the study [15] conducted a combination of SMOTE, ENN, and TomekLinks against SVM can improve accuracy performance from SVM of 2% to 23% on SMOTE-ENN technique.

Based on the background described above and previous research, some problems can be formulated, firstly how to compare the performance of the classification model with the C4.5, KNN, and Logistic regression algorithms by performing hyperparameter tuning using GridSearchCV and RandomizedSearchCV, secondly how to overcome dataset imbalance using SMOTE and TomekLinks and using ensemble learning, namely voting classifier. This study aims to create a prediction model for cardiovascular heart disease patients using the C4.5 algorithm, KNN, logistic regression, and the ensemble learning voting classifier method. it is hoped that this research can help researchers in the field of heart disease in choosing the right model or method for their research and can help improve the services of medical experts.

Method

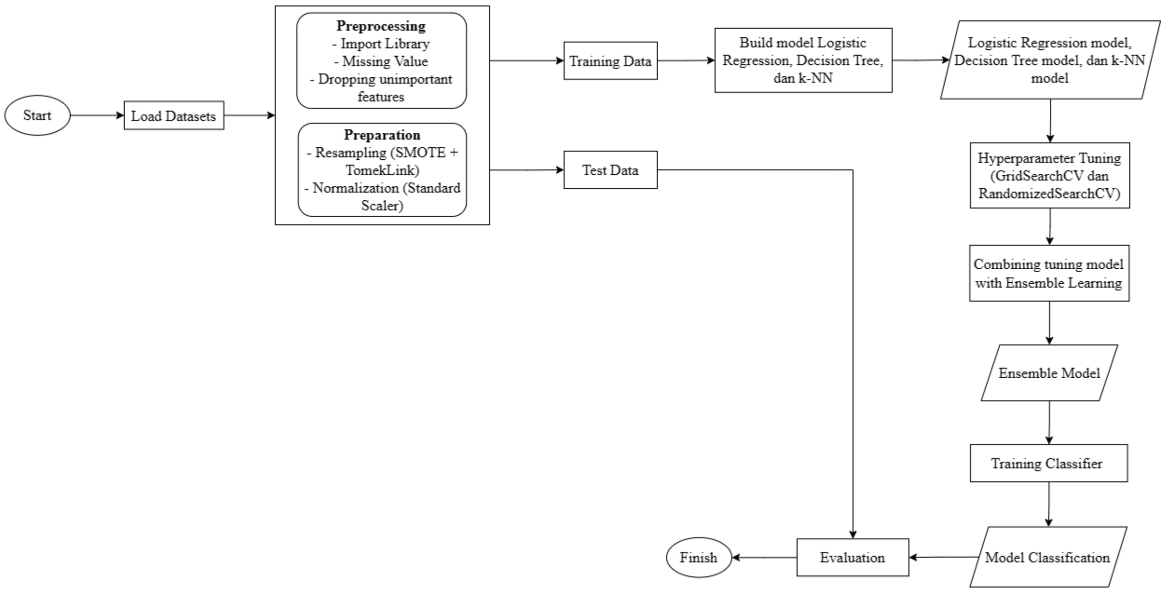

Figure 1. Research Flow

In the system design process, there were several steps that must be carried out. In the process flow of making predictions shown in **Figure 1**, it started with preparing the dataset, preprocessing, and model evaluation.

A. Data Collection and Exploration

This study used a dataset containing information about the data records of patients with heart failure. This dataset had dimensions of 299 rows and 13 columns with all the numeric/continuous type features/columns obtained from https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data. This dataset has been widely used by researchers, one of them [1] was an early researcher and contributor to the dataset. The dataset consisted of several independent variables and one dependent/target variable (DEATH_EVENT). Independent variables/predictors included age, anemia, creatine phosphokinase, diabetes, ejection fraction, high blood pressure, and so on.

B. Preprocessing

Before dividing the data into training data and testing data, the data went through a preprocessing process first. The first step was to eliminate or fill in missing values, to delete unused columns, and to delete outliers. The delated *values* were those with unique characteristics or beyond the limit compared to other data so that if maintained it would affect model performance. Deleting outliers was only applied to features that were numeric [11].

C. Resampling

In this study, resampling would be carried out where the imbalance data was oversampled using SMOTE, then the data would be reduced by the under-sampling method using Tomek-Links. The following is a brief explanation of the SMOTE and tomek-links methods that was used in this study.

1. SMOTE
   The resampling technique used Synthetic Minority Oversampling Technique (SMOTE) to overcome the imbalance of classes 0 and 1 in the DEATH_EVENT column. SMOTE worked by elevating minority class data so that it was balanced with majority class data [6][8][16].

2. Tomek-Link
   The tomek-links technique was a method that works by subtracting samples (under-sampling) utilizing 2 data that are close to each other between the minority class and the majority class [15].

D. Data Preparation

Row data values varied widely and could cause bias in the model training. The simplest method to solve this problem was to use feature scaling. It was a technique used to place data in a denser range. Feature scaling was the most important step in machine learning before creating a model [17]. In this experiment, researchers used Standard Scaler for the scaling process available in the scikit-learn framework.

E. Hyperparameter Tuning

Hyperparameter tuning is a method that can be used to optimize machine learning algorithms optimally [17]. This study used grid search and randomized search techniques to find the optimal hyperparameter value in the model.

1. Grid Search Cross Validation
   Grid search cross validation is a method used to implement grid search and cross-validation technology, where the method works by combining models and hyperparameters by selecting only one of each and validating for each combination [17].

2. Randomized Search Cross Validation
   Randomized search cross validation is a technique that works by performing random analysis according to predetermined and stable hyperparameters after looping according to the desired iteration [11].

F. Building Algorithm Model

Creating a model for this classification system uses predetermined algorithms, namely logistic regression, k-NN, and C4.5. After creating all three models and all models had been tuned, the models would be combined using the ensemble learning method, namely the voting classifier technique to get the majority value of the three algorithms.

1. K-Nearest Neighbor
   K-Nearest Neighbor (KNN) is a classification algorithm for learning data or objects based on the value of the k-nearest neighbor [16][18]. Where k serves as the distance or similarity between the data [19]. The algorithm works by finding the sample value closest to the sample input k and calculating the result based on the largest sample size of the input k [18].

2. Logistic Regression
   In the context of regression and classification problems, logistic regression is a technique that fits the category of supervised learning. In the case of classification, logistic regression uses probabilities to make predictions based on categorical data [20].

3. C4.5 (Decision Tree)
   The C4.5 algorithm, also known as the decision tree, is a classifiable algorithm with fast processing speed that can handle numerical and discrete data as well as identify and remove already missing attributes and generate easy-to-implement rules [21]. The C4.5 algorithm uses an attribute as *root*, creating a branch for each value in the case, and the process will continue to be repeated until each branch case has the same class [22].

4. Ensemble Learning – Voting Classifier
   Ensemble learning is the process by which a single algorithm learns from data by combining several different algorithms or models to provide results with higher accuracy. Voting, bagging, boosting, and stacking are some of the ensemble learning techniques that can be used [18]. In this study, the authors used the voting classifier to get the majority value of the predetermined algorithm [23] by combining all prediction models [24].

Results and Discussion

In the preprocessing stage, it produced clean data by eliminating non-essential features and removing or filling in empty values, then generating balanced data using resampling techniques, and then getting an improved model with hyperparameter tuning and ensemble learning methods. Testing stage was carried out to determine the performance of all models of the algorithm. The best-performing model can be used to classify the survival rate of heart disease patients.

A. Datasets

After preprocessing, the dataset used had dimensions of 299 rows and 10 columns, with all features being of numeric type. Datasets were public data obtained from (https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data). The dataset was divided into two parts, namely training data and testing data. The training data was used to train algorithms in the formation of a model, while testing data was used to measure the performance obtained from the model training process with training data. In this study, we divided the data with a ratio of 90:10, 80:20, and 70:30 in the training and testing data. Then the performance of each data ratio was tested. An example of the data used can be seen in **Table 1**.

Table 1. Datasets

No	Age	Anemia	Creatine Phosphokinase	Ejection Fraction	High Blood Pressure	Platelets	Serum Creatinine	Serum Sodium	Time	Death Event
1	75	0	582	20	1	265000	1.9	130	4	1
2	55	0	7861	38	0	263358	1.1	136	6	1
3	65	0	146	20	0	162000	1.3	129	7	1
4	50	1	111	20	0	210000	1.9	137	7	1
5	65	1	160	20	0	327000	2.7	116	8	1
…	…	…	…	…	…	…	…	…	…	…
295	62	0	61	38	1	155000	1.1	143	270	0
296	55	0	1820	38	0	270000	1.2	139	271	0
297	45	0	2060	60	0	742000	0.8	138	278	0
298	45	0	2413	38	0	140000	1.4	140	280	0
299	50	0	196	45	0	395000	1.6	136	285	0

B. Evaluation

To find the level of performance, accuracy, recall, precision, and FI score were calculated using a formula based on the following **Table 2**.

Table 2. Confusion Matrix

Actual Class		Predicted Class	
		Positive	***Negatives***
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Accuracy is the result value for determining the accuracy of the model [25]. Accuracy is the comparison of the amount of true and negative positive data with the overall data [3][15]. The accuracy formula can be seen in Equation 1.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Precision was a comparison of positive true predictions divided by overall positive predicted results [3][15]. The formula of precision can be seen in Equation 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall or sensitivity was the ratio of positive correct predictions to overall positive correct data [3][15]. The formula of the recall can be seen in Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1 Score was a weighted average comparison of precision and recall [15][25]. The formula of the F1 score can be seen in Equation 4.

$$F1\ Score = 2\frac{Recall\ \times\ Precision}{Recall\ +\ Precision} \tag{4}$$

In the testing phase, the data was divided into two, namely train data and test data with three scenarios, namely with a ratio of 90:10, 80:20, and 70:30. Once the data was divided according to the specified proportions, it would be normalized with the Standard Scaler. After that, the data would be trained with a predetermined model, namely Logistic Regression, KNN, C4.5, and combined with a voting classifier. Then accuracy, recall, F1 score, precision, and ROC/AUC were evaluated to search for the performance of the model. Below are the performance results of each model of the algorithm.

Table 3. Evaluation Dataset Ratio 90:10

Algorithm Models	Accuracy	Recall	Precision	F1 Score	ROC/AUC
Logistic Regression	80.55%	86.48%	78.05%	82.05%	80.39%
KNN	88.89%	94.6%	85.37%	89.74%	88.73%
C4.5	77.78%	78.38%	78.38%	78.38%	77.76%
Voting Classifier	88.89%	97.3%	83.72%	90%	88.65%

In **Table 3** of the dataset divided by a ratio of 90:10, the highest accuracy was by voting classifier and KNN with a value, 88.89%, logistic regression obtained an accuracy value, 80.55%, the C4.5 model got the smallest accuracy value, 77.78%. Then the highest ROC / AUC value was obtained by KNN with a value of 88.73% so that it was included in the good classification, as well as voting classifier and logistic regression because each ROC / AUC value is in the range of 0.80-0.90. While the smallest ROC/AUC value by the Decision Tree with a value of 77.76% was included in the fair classification because it was in the range of 0.70-0.80 [26].

Table 4. Evaluation of 80:20 Ratio

Algorithm Models	Accuracy	Recall	Precision	F1 Score	ROC/AUC
Logistic Regression	75%	73.78%	67.86%	70.37%	74.7%
KNN	84.38%	88.46%	76.67%	82.14%	85.02%
C4.5	81.25%	73.08%	79.17%	76%	79.96%
Voting Classifier	87.5%	96.15%	78.13%	86.21%	88.86%

In **Table 4** of the dataset divided by a ratio of 80:20, the highest accuracy was by the voting classifier with a value of 87.5%, logistic regression obtained an accuracy value of 75%, KNN experienced a decrease in accuracy compared to the previous one which was 84.38%, and the C4.5 model experienced an increase in accuracy which was 81.25%. Then the highest ROC / AUC value by the voting classifier with a value of 88.86% experienced an increase of 0.21% and it was still included in the good classification because the value of 88.86% was in the range of 0.80-0.90 [26]. Then the ROC/AUC value of the C4.5 model increased by 2.2%, while the ROC/AUC KNN and Logistic Regression values decreased, so the model was included in the fair classification.

Table 5. Evaluation of 70:30 Ratio

Algorithm Models	Accuracy	Recall	Precision	F1 Score	ROC/AUC
Logistic Regression	89.47%	89.65%	89.65%	89.65%	89.77%
KNN	91.23%	89.65%	92.86%	91.23%	91.26%
C4.5	89.47%	89.65%	89.65%	89.65%	89.47%
Voting Classifier	94.74%	93.1%	96.42%	94.74%	94.77%

In **Table 5** of the dataset divided by a ratio of 70:30, there was a significant increase in all evaluation schemes. The highest accuracy was shown by voting classifier with a value of 97.74% and the ROC / AUC value was almost the same, 94.77%, as well as in all models it had an accuracy value that was almost the same as the ROC / AUC value. There were two models whose ROC/AUC values were included in the excellent classification, namely voting classifier, and

KNN because the values were in the range of 0.90-1.00 [26]. While the other two models were included in the good classification.

In the next performance test, it can be seen the score accuracy value in the training process with training and testing data to evaluate the occurrence of overfitting and underfitting in all models built. The results of performance testing can be seen in **Table 6**, **Table 7**, and **Table 8** as follows.

Table 6. Accuracy score of 90:10 Ratio

Algorithm Model	Training set score	Testing set score
Logistic Regression	0.8984	0.9158
KNN	1.0	0.9127
C4.5	0.9265	0.8575
Voting Classifier	0.9686	0.8889

In the test results listed in **Table 6** with a ratio of 90:10, KNN obtained the highest training score of 100% and a lowest testing score at 91.27%. The gap was 9% just as other models had an average gap of 4-9%. Logistic regression had the smallest training value of the three models, so the model underfitting where accuracy training was smaller than testing [18] and C4.5 obtained the smallest testing value of the three models.

Table 7. Accuracy score of 80:20 Ratio

Algorithm Model	Training set score	Testing set score
Logistic Regression	0.9257	0.8432
KNN	1.0	0.8957
C4.5	0.9372	0.9059
Voting Classifier	0.9685	0.875

In the test results listed in **Table 7** with a ratio of 80:20, the results of several models increased significantly and there was no underfitting. The highest accuracy training was still obtained by KNN and the smallest accuracy testing shown by logistic regression decreased by about 7% compared to the previous results. The other two models also experienced a decrease in accuracy testing of about 1.5% compared to the first test. The C4.5 model experienced an increase in accuracy testing of about 4.8% compared to the first test result.

Table 8. Accuracy score of 70:30 Ratio

Algorithm Model	Training set score	Testing set score
Logistic Regression	0.8995	0.9692
KNN	1.0	0.9655
C4.5	0.9030	0.8855
Voting Classifier	0.9598	0.9474

In the test results listed in **Table 8** with a ratio of 70:30, the changes in the results obtained were quite balanced in several models. The highest accuracy testing was still obtained by KNN, and in this test, logistic regression underfitting with a distance difference of 6.97%. Then the voting classifier model had almost the same accuracy training and testing values that only differed by 1%, so there was a good decrease in overfitting.

Visualization of the performance results of the testing process between training and testing with data sharing of 90:10, 80:20, and 70:30 can be seen in **Figure 2**, **Figure 3**, and **Figure 4**.

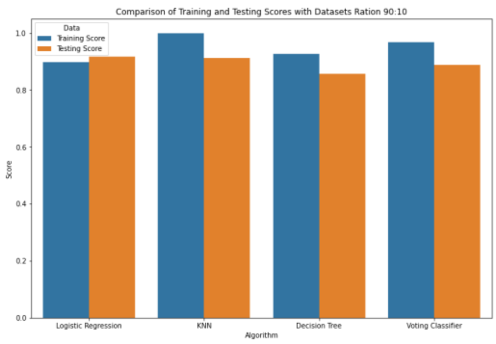

Figure 2. Visualization of Accuracy Score of Ratio 90:10

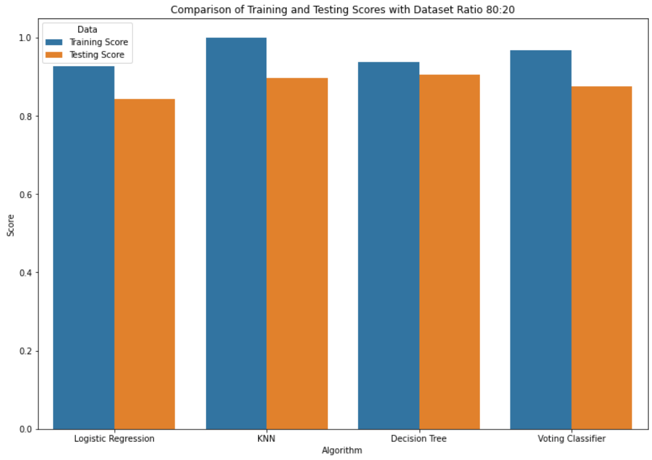

Figure 3. Visualization of Accuracy Score of Ratio 80:20

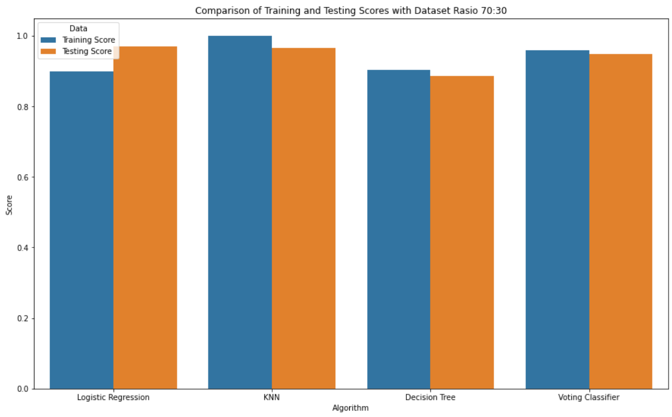

Figure 4. Visualization of Accuracy Score of Ratio 70:30

Based on the visualizations in **Figure 2**, **Figure 3**, And **Figure 4**, the performance of the voting classifier model looked quite prominent and showed stable performance at a ratio of 70:30 whereas KNN obtained the highest performance from all algorithm models. However, the performance result was not good enough compared to voting classifier. Then the Logistic Regression model indicated that its performance was not good, because of the ratio of 90:10 and 70:30 underfitting.

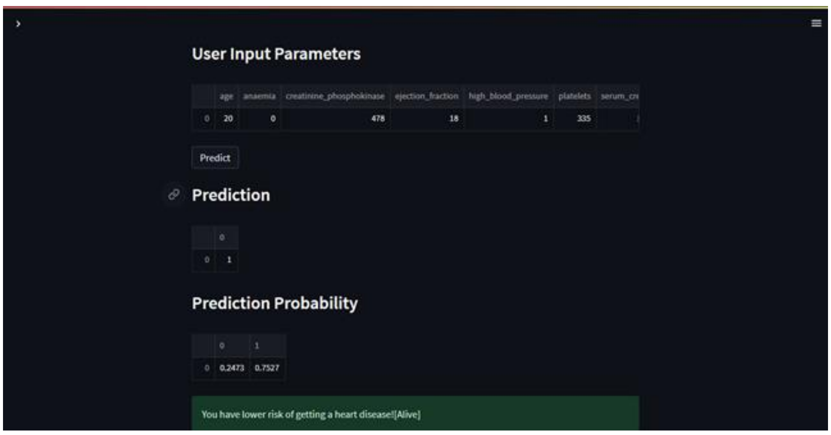

Figure 5. Result of Voting Classifier

The implementation of the created system can be seen in **Figure 5** generated using the voting classifier model. The program run well and could successfully classify based on new data entered by the user by generating an 'Alive' answer.

Conclusion

The current research has been working on improving the prediction performance of the patient's life expectancy of cardiovascular heart failure by implementing logistic regression, KNN, C4.5 algorithms, the ensemble learning method of voting classifier, and the techniques to address data imbalance and hyperparameter tuning. The conclusion drawn from this research is that the accuracy of the voting classifier obtains a higher value compared to KNN, logistic regression, and C4.5, using datasets in the ratios 80:20 and 70:30. The accuracy was 94.74% and 87.5% respectively. The voting classifier demonstrates stable performance at the 70:30 ratio, with an accuracy difference between training and testing scores below 1%. This indicates that using the voting classifier is more beneficial in terms of learning from data and enhancing the algorithm's performance in classifying the patient's life expectancy of cardiovascular heart failure into 'Survived' or 'Deceased', compared to logistic regression, KNN, and C4.5.

Acknowledgment

Researchers would like to thank Amikom Purwokerto University for the support and opportunities provided to researchers through the MSIB Program of the Ministry of Education and Culture, Research and Technology 2022, both materially and non-materially. Researchers are also grateful to the supervisors who always provide me a genuine and meaningful support, advice, and input to the researcher so that this research can be completed on time.

The next research can conduct experiments with other algorithms that can be used in ensemble learning such as Random Forest, Gradient Boosting, or Support Vector Machine. Then, it can try other ensemble learning techniques like hard voting, bagging, or stacking.

References

[1] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–16, 2020.

[2] Edric and S. P. Tamba, "Prediksi penyakit gagal jantung dengan menggunakan random forest," *JUSIKOM PRIMA (Jurnal Sist. Inf. dan Ilmu Komput. Prima)*, vol. 5, no. 2, pp. 176–181, 2022.

[3] D. A. M. Reza, A. M. Siregar, and Rahmat, "Penerapan Algoritma K-Nearest Neighbord Untuk Prediksi Kematian Akibat Penyakit Gagal Jantung," *Sci. Student J. Information, Techbnology Sci.*, vol. 3, no. 1, pp. 105–112, 2022.

[4] W. H. Organization, "Cardiovascular Diseases," *World Health Organization*, 2023. https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 (accessed Jan. 19, 2023).

[5] S. E. Awan, M. Bennamoun, F. S. Id, F. M. Sanfilippo, B. J. Chow, and G. D. Id, "Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death," *PLoS One*, vol. 14, no. 6, pp. 1–13, 2019.

[6] D. Mualfah *et al.*, "Jurnal Computer Science and Information Technology (CoSciTech) algoritma random forest," *CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 2, pp. 107–113, 2022, doi: https://doi.org/10.37859/coscitech.v3i2.3912.

[7] L. Budianti and Suliadi, "Metode Weighted Random Forest dalam Klasifikasi Prediksi," in *Bandung Conference Series: Statistics*, 2017, pp. 103–110,doi: https://doi.org/10.29313/bcss.v2i2.3318.

[8] S. M. Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, "Penerapan Synthetic Minority Oversampling Technique (SMOTE) untuk Imbalance Class pada Data Text Menggunakan KNN," *Syntax J. Inform.*, vol. 11, no. 02, pp. 56–67, 2022.

[9] D. S. Permana and A. Silvanie, "Prediksi Penyakit Jantung menggunakan Support Vector Machine dan Python pada Basis Data Pasien di Cleveland," *Junif J. Nas. Inform.*, vol. 2, no. 1, pp. 29–34, 2021.

[10] W. Nugraha, "Prediksi Penyakit Jantung Cardiovascular menggunakan Model Algoritma Klasifikasi," *J. Manag. dan Inform.*, vol. 9, no. 2, pp. 78–84, 2021.

[11] M. R. Mubarok *et al.*, "Hyper-Parameter Tuning Pada Xgboost Untuk Prediksi Keberlangsungan Hidup Pasien Gagal," *Kumpul. J. Ilmu Komput.*, vol. 09, no. 02, pp. 391–401, 2022.

[12] Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 298–306, 2022.

[13] S. Adi and A. Wintarti, "Komparasi Metode Support Vector Machine (SVM), K-Nearest Neighbors (KNN),

dan Random Forest (RF) untuk Prediksi Penyakit Gagal Jantung," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 02, pp. 258–268, 2022.

[14] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree , Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *J. Inform.*, vol. 4221, no. 18, pp. 239–248, 2022.

[15] A. Indrawati, "Penerapan Teknik Kombinasi Oversampling dan Undersampling untuk Mengatasi Permasalahan Imbalanced Dataset," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 1, pp. 38–43, 2021, doi: 10.33387/jiko.

[16] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor," *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.

[17] W. Nugraha and A. Sasongko, "Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search," *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022.

[18] M. Ma, A. Prayogo, P. Subarkah, and F. Nida, "Sentiment analysis of customer satisfaction levels on smartphone products using Ensemble Learning," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 339–347, 2022, doi: http://dx.doi.org/10.33096/ilkom.v14i3.1377.339-347.

[19] A. Alhamad, A. I. S. Azis, B. Santoso, and S. Taliki, "Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble – Weighted Vote," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 352–360, 2019.

[20] F. Handayani *et al.*, "Komparasi Support Vector Machine , Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 7, no. 3, pp. 329–334, 2021.

[21] H. Hasanah and Nurmalitasari, "Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung," in *Prosiding Seminar Nasional Teknologi dan Sains*, 2023, vol. 2, pp. 13–18.

[22] P. D. Putra and D. P. Rini, "Prediksi Penyakit Jantung dengan Algoritma Klasifikasi," in *Prosiding Annual Research Seminar*, 2019, vol. 5, no. 1, pp. 95–99.

[23] M. Amine, S. El, and E. L. Mohammed, "Breast Cancer Prediction and Diagnosis through a New Approach Breast Cancer Prediction and Diagnosis through a New Approach based on Majority Voting Ensemble Classifier based on Majority Voting Ensemble Classifier," in *Procedia Computer Science*, 2021, vol. 191, pp. 481–486, doi: 10.1016/j.procs.2021.07.061.

[24] A. C. Lagman, L. P. Alfonso, M. L. I. Goh, J. P. Lalata, and J. P. H. Magcuyao, "Classification Algorithm Accuracy Improvement for Student Graduation Prediction Using Ensemble Model," *Int. J. Inf. Educ. Technol.*, vol. 10, no. 10, pp. 723–727, 2020, doi: 10.18178/ijiet.2020.10.10.1449.

[25] D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.

[26] A. Merdekawati, "Komparasi Algoritma Data Mining dan Perancangan Aplikasi Prediksi Harapan Hidup Pasien Gagal Jantung," *CSRID J.*, vol. 14, no. 3, pp. 188–202, 2022, https://csridjournal.potensi-utama.ac.id/index.php/CSRIDjournal/article/download/3/1.