

EXPLORATORY DATA ANALYSIS OF CLINICAL HEART FAILURE USING A SUPPORT VECTOR MACHINE

*Putri Tua Sinaga^{*1}, Saldia Sari Purba², David Wiranto³, Okta Jaya Maharaja⁴, Evta Indra⁵*
^{1,2,3,4,5}Indonesian Prima University
Jl. Sampul No. 3, Sei Putih Bar., Kec. Medan Petisah, Medan City, North Sumatra 20118
Email : putrisinaga22@gmail.com*

ABSTRACT-This study aims to explore the clinical data of patients diagnosed with heart failure using the Support Vector Machine (SVM) algorithm as a classification method. Clinical data from verified patients has been collected and analyzed to identify patterns, associations, and risk factors contributing to heart failure risk. The exploratory data analysis results reveal essential clinical data characteristics and provide initial insight into patient profiles and clinical variables that can influence heart failure risk. The SVM model was built to predict the risk of heart failure based on clinical data. This model is evaluated using classification metrics such as F1-Score and accuracy. Evaluation results show good performance with an F1-Score reaching 0.83, which indicates a reasonable degree of accuracy and balance in predicting the risk of heart failure. The conclusion of this study shows the potential of the classification model as a tool in managing heart failure patients. This model can help medical personnel identify high-risk patients and provide appropriate treatment to prevent disease progression and improve prognosis. However, these results need further verification with more in-depth analysis and validation using broader data. This model can help medical personnel identify high-risk patients and provide appropriate treatment to prevent disease progression and improve prognosis. However, these results need further verification with more in-depth analysis and validation using broader data. This model can help medical personnel identify high-risk patients and provide appropriate treatment to prevent disease progression and improve prognosis. However, these results need further verification with more in-depth analysis and validation using broader data.

Keywords: Exploratory Data Analysis, Heart Failure, Classification, Python, Support Vector Machine

1. INTRODUCTION

The heart is a vital organ that plays a role in pumping and circulating blood throughout the body so that the organs and tissues of the body can operate correctly. However, several factors can cause disturbances in the heart, which result in this organ not being able to function optimally or what is commonly known as heart failure [1], [2]. Heart disease is considered the most dangerous disease in human life worldwide. Specifically, in this type of disease, the heart cannot pump the required amount of blood to the remaining organs of the human body to achieve normal function [3]. Various factors, such as high blood pressure, coronary heart disease, heart valve problems, and others, cause heart failure. Heart failure is one of the leading causes of death in the world.

To improve the treatment and management of patients with heart failure, exploratory data analysis is the key to understanding relationships and patterns in clinical data. EDA is a technique to explore data sets to extract valuable and actionable information, identify relationships among explanatory variables, detect errors, and select appropriate models [4]. It uses descriptive statistics and graphical tools to develop an understanding of the data [4], [5]. EDA is used mainly

to maximize insights into the dataset, detect outliers and anomalies, and test the underlying assumptions [4], [6] to support clinical decision-making. One helpful tool for classifying and predicting clinical data is the Support Vector Machine (SVM). SVM is one of the famous and influential machine learning techniques that can be used for classification and regression. The support vector machine (SVM) method is a machine learning method that has a structural risk minimization principle that aims to find the best hyperplane that separates two groups in the input space [7].

SVM can classify heart failure patients into two classes, for example, patients with low and high risk. Previous studies that tested heart failure as a reference for researchers are "Gaussian Naïve Bayes Algorithm Analysis of Data Classification of Heart Failure Patients" [8], aimed at classifying data on patients suffering from heart failure. She talked about using the naïve Bayes algorithm to categorize diseases affected by health conditions that show increased heart failure rates with a population of one hundred individuals. As a result, their level of welfare is significantly affected. Congestive heart failure can be classified based on the level of physical ability and the duration of the condition. A study by [9] investigated the factors associated with a patient's quality of life suffering from congestive heart failure. It focused on the level of physical ability and the length of illness. Research conducted by [10] analyzes the relationship between certain factors and heart disease using the KNN method and cross-validation, which aims to identify risk factors contributing to the disease's development. By using KNN, patient data with similar characteristics can be grouped.

Meanwhile, cross-validation creates an accurate model by dividing the data into several subsets and testing them using others. Putri hopes to provide valuable information on managing and preventing heart disease by using these two methods. Cross-validation creates an accurate model by dividing the data into several subsets and testing them using other subsets. Putri hopes to provide valuable information on managing and preventing heart disease by using these two methods. Cross-validation creates an accurate model by dividing the data into several subsets and testing them using other subsets. Putri hopes to provide valuable information on managing and preventing heart disease by using these two methods.

In conducting exploratory data analysis of clinical data of heart failure patients and implementing SVM as a classification tool to assist in diagnosing and predicting the risk of heart failure. This research is expected to reveal new insights, support more informed clinical decision-making, and improve the management of patients suffering from heart failure.

2. RESEARCH METHODS

2.1 Research Methods

Research methodology's importance in ensuring that the data collected is relevant, valid, and reliable. In addition, research methodology also helps ensure that research is carried out systematically and objectively so that the results can be accounted for and used to make better decisions.

This study uses the Google Collaboratory with the Python programming language in Exploratory Data Analysis and implementation of the Support Vector Machine classification method. And the framework for this research can be seen in Figure 1.

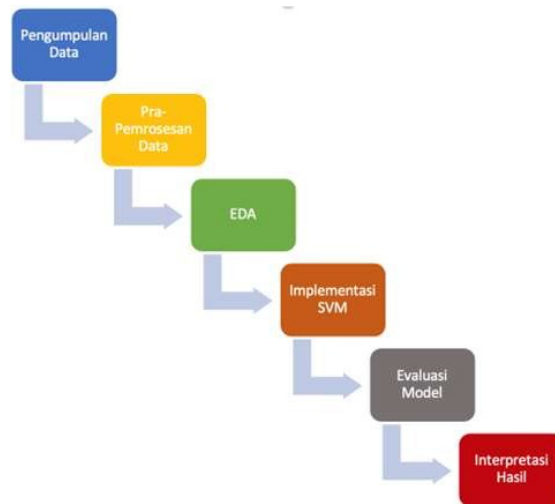


Figure 1 Framework

2.2 Data collection

Data Acquisition is retrieving data from valid sources before the data is processed and used to create Machine Learning models [11]. This first step is to collect clinical data of patients with heart failure diagnoses from relevant and reliable sources. The data collected should include clinically relevant variables, such as age, sex, blood pressure, cholesterol levels, cardiac function, and other variables related to heart failure. Where the dataset was obtained from the Rithik Khota repository Kaggle site "Heart Failure Clinical Records Classification" [12]. The dataset has 13 columns and 299 rows of data which can be seen in more detail in Figure 2.

Heart_failure_clinical_records_dataset.csv (12.54 kB)

Detail	Compact	Column
# age	# anaemia	# creatinine...
# diabetes	# ejection_fr...	# high_bloo...
# platelets	# serum_cre...	# serum_so...
# sex		

Image Error! No text of specified style in the document
 Datasets

2.3 Preprocessing Data

The preprocessing phase prepares suitable data for building and training Machine Learning models [11]. After the data is collected, preprocessing is carried out to clean the data from missing or invalid values, handle outliers, and normalize variables if necessary. This step is essential to ensure that the data used in the analysis is of good quality and ready to be processed.

2.4 Exploratory Data Analysis (EDA)

Data analysis is a method used to find out how to describe data, data relationships, data semantics, and data limitations that exist in an information system [13], [14]. The main task of EDA is to clean data, describe data, see data distribution, compare relationships between data and conclude data [15]–[17]. After the data was prepared, an exploratory analysis was carried out to understand the characteristics of the data, the distribution of variables, and the relationship between clinical variables. Visualizations such as graphs, charts, and plots will help identify exciting patterns and initial insights.

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a Machine Learning algorithm that is included in the Supervised Learning model or supervised learning related to data analysis and pattern recognition [11] and is a suitable method for classification and regression problems, both linear and nonlinear [18]. The SVM concept can be expressed as an attempt to find an optimal classifier/hyperplane function that can separate two data sets from two different classes [7]. SVM is a linear classifier that works for binary classification but was later extended to perform multilayer classification using techniques such as one vs. one, one for all and directed acyclic graphs [19], [20]. Furthermore, to be able to work on nonlinear problems by bringing core concepts into high workspaces.

After understanding the data, the Support Vector Machine (SVM) algorithm is implemented as a classification method. Clinical data will be divided into training data and testing data. The SVM model will be trained with training data and tested on test data to measure its performance.

2.6 Model Evaluation

The SVM model that has been built will be evaluated using classification metrics such as accuracy, precision, recall, and F1-score. F1-Score is a metric that measures the balance between the precision and memory of a classification model. F1-Score ranges from 0 to 1, where 1 indicates perfect performance, and 0 indicates poor performance. Accuracy is a metric that measures how accurately a model predicts the correct class. Accuracy ranges from 0 to 1, where 1 indicates a perfect prediction, and 0 indicates a poor prognosis. Support is the number of samples in the test set that support each class. High F1-Score and Accuracy results show that your SVM model has produced good results.

In addition to considering the F1-Score and Accuracy values, it is also essential to understand the research context, objectives, and sample size to ensure proper interpretation of the model evaluation results. The larger the sample size and the more balanced the class distribution of the test data, the more reliable the model evaluation results will be. This evaluation will provide an overview of the extent to which the model can predict the risk of heart failure based on clinical data.

2.7 Interpretation of Results

The results from the exploratory analysis of the data and evaluation of the SVM model will be interpreted to gain valuable insights into the model's clinical relationship, risk factors, and predictive ability. This interpretation will provide a detailed explanation of the research results and their implications in the treatment and management of heart failure patients.

3. RESULTS AND DISCUSSION

3.1 Exploratory Data Analysis (EDA)

In this study, data manipulation was carried out, namely changing the 'platelets' column to 'platelets' to make it easier to read the dataset. More details can be seen in Figure 3.

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	trombosit	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7	1
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8	1

**Image Error! No text of specified style in the document.
Column name manipulation**

The dataset used has a description like in Figure 4, where the minimum, middle, quartile, and maximum values can be identified. Next, check for duplicate data and null data in the dataset. However, there were no same or invalid data in this research dataset.

	count	mean	std	min	25%	50%	75%	max
age	299.0	60.833893	11.894809	40.0	51.0	60.0	70.0	95.0
anaemia	299.0	0.431438	0.496107	0.0	0.0	0.0	1.0	1.0
creatinine_phosphokinase	299.0	581.839465	970.287881	23.0	116.5	250.0	582.0	7861.0
diabetes	299.0	0.418060	0.494067	0.0	0.0	0.0	1.0	1.0
ejection_fraction	299.0	38.083612	11.834841	14.0	30.0	38.0	45.0	80.0
high_blood_pressure	299.0	0.351171	0.478136	0.0	0.0	0.0	1.0	1.0
trombosit	299.0	263358.029264	97804.236869	25100.0	212500.0	262000.0	303500.0	850000.0
serum_creatinine	299.0	1.393880	1.034510	0.5	0.9	1.1	1.4	9.4
serum_sodium	299.0	136.625418	4.412477	113.0	134.0	137.0	140.0	148.0
sex	299.0	0.648829	0.478136	0.0	0.0	1.0	1.0	1.0
smoking	299.0	0.321070	0.467670	0.0	0.0	0.0	1.0	1.0
time	299.0	130.260870	77.614208	4.0	73.0	115.0	203.0	285.0
DEATH_EVENT	299.0	0.321070	0.467670	0.0	0.0	0.0	1.0	1.0

Figure 4 Description of the dataset

3.2 Outlier Detection

Because too many outlier data were detected in this research dataset, the outliers were ignored. Outlier data can be seen more clearly in Figure 5.

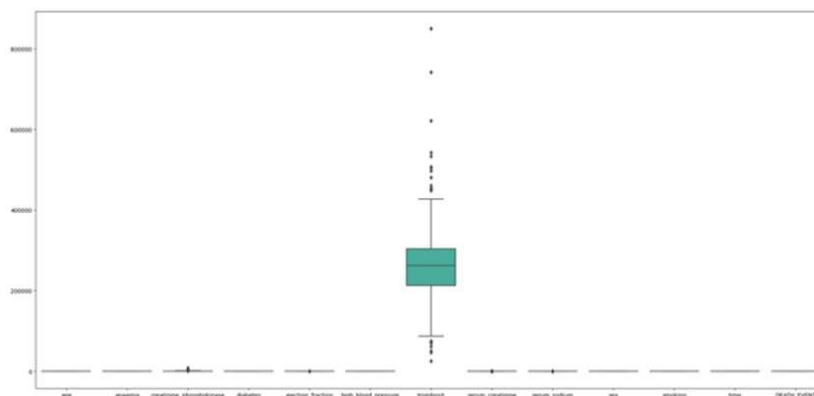


Figure 5 Outliers

3.3 Setting EDA Targets

In this study, death is the target. The dataset shows that the death rate is lower than the survival rate, which can be seen in Figure 6. Next, we will analyze the variables that influence the death rate in cases of heart failure.

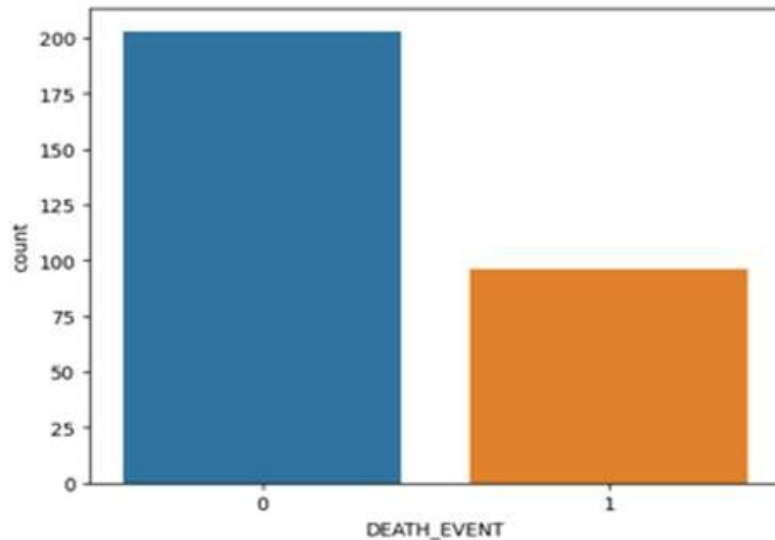


Figure 6 Death data

3.4 Univariate Analysis

In Figure 7 you can see the data distribution in the dataset using a histogram which aims to make it easier to read heart failure patient data. The data is in the form of age diagrams, creatinine phosphokinase, ejection fraction, platelets, and time. The histogram shows that the highest number of heart failure patients is at 60 years, and the lowest is at 70 years and over.

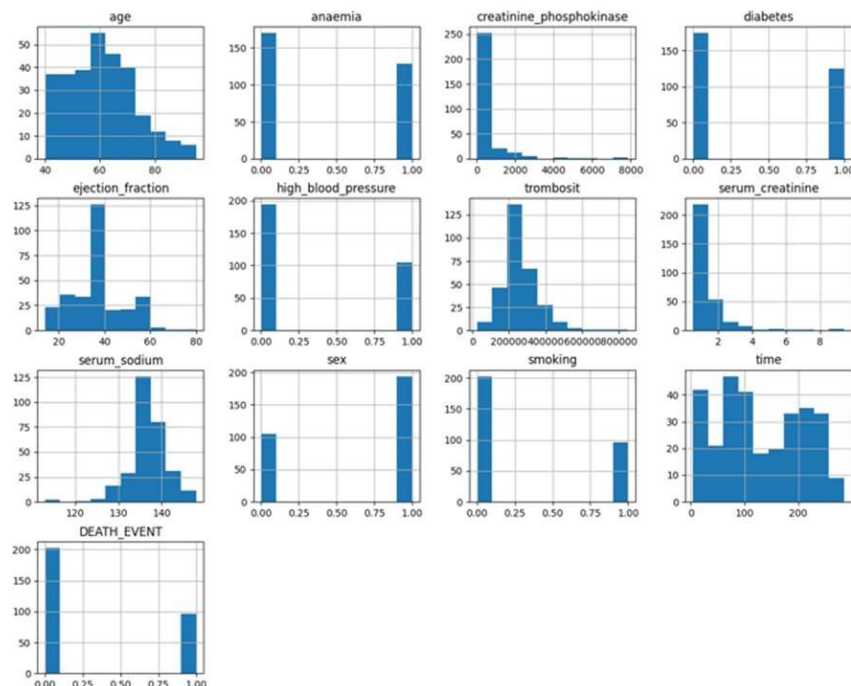


Figure 7. Visualization of univariate analysis

3.5 Multivariate Analysis

The multivariate analysis showed that the mortality rate due to heart failure in patients with anemia, diabetes, and high blood pressure was slightly lower. The death rate from heart failure is more common in males than females. In addition, patients who do not smoke have a higher risk of heart failure. Figure 8 presents a visualization of the interrelationship of variables in heart failure cases.

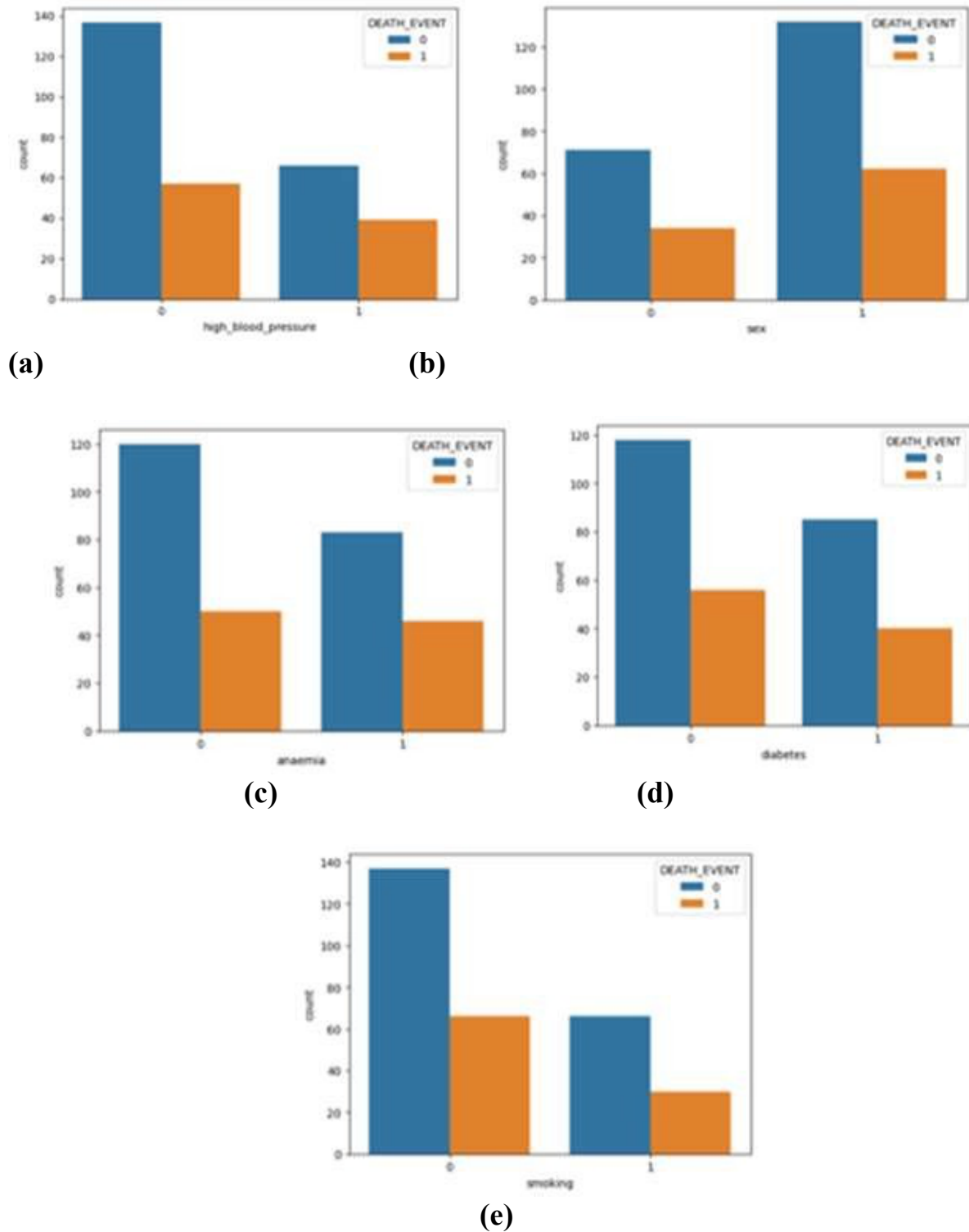


Figure 8 Visualization of multivariate analysis (anemia vs. death event (a), diabetes vs. death event (b), high blood pressure vs. death event (c), sex vs. death event (d), smoking vs. death event (e))

To find out the relationship or correlation between variables, visualize it using a heatmap, where values close to number 1 have a high correlation and numbers close to -1 have a low correlation. In contrast, a value of 0 has a neutral correlation. In Figure 9 it can be

seen that the time and death event variables have a low correlation, while the sex and smoking variables have a high correlation.

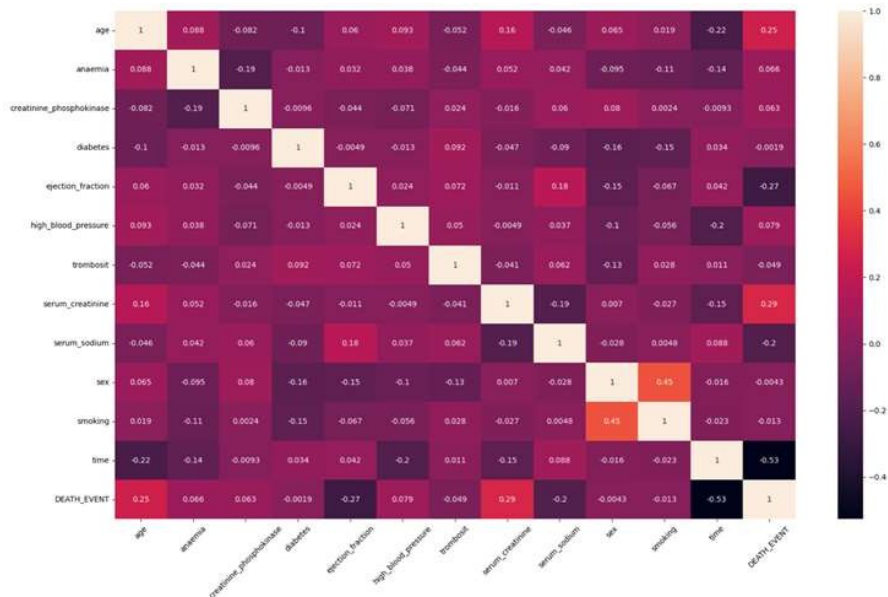


Figure 9 Correlation of heart failure disease variables

3.6 Implementation of Support Vector Machine (SVM)

Before model building, data splitting/partitioning is carried out, namely by dropping the death event variable and making this variable Y, while the other variables become the X axis. Next, train and test the dataset using the sklearn python library. More details can be seen in Figure 10.

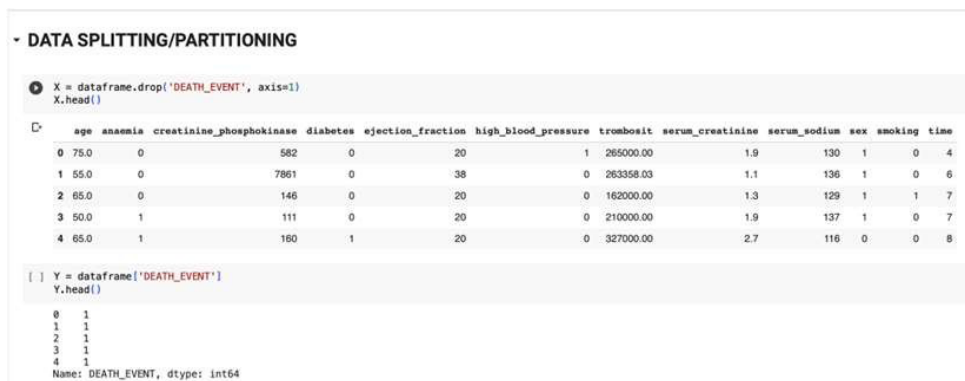


Figure 10 Drop variable death event.

In Y training resulted in 158 survivors alive and 81 died, while the Y testing resulted in 45 surviving or alive and 15 dead, as shown in Figure 11.


```

[57] from sklearn import preprocessing
    col_names = list(X.columns)
    scaler = preprocessing.StandardScaler()
    X_scaled= scaler.fit_transform(X)
    X_scaled= pd.DataFrame(X_scaled, columns=col_names)

[60] from sklearn.model_selection import train_test_split
    from sklearn.metrics import accuracy_score

[59] # MEMBUAT FUNGSI PARTITIONING/SPLITTING
    X_train, X_test, Y_train, Y_test = train_test_split(X_scaled, Y, test_size=0.2,
                                                    random_state=101, shuffle=True)

[61] # MELIHAT DATASET TRAINING
    Y_train.value_counts()

0    158
1     81
Name: DEATH_EVENT, dtype: int64

[62] # MELIHAT DATASET TESTING
    Y_test.value_counts()

0     45
1     15
Name: DEATH_EVENT, dtype: int64
    
```

Figure 11 Data Splitting/partitioning

In the Train model and Support Vector Machine (SVM) algorithm test that has been made, the resulting training accuracy value is 88.70%, and the testing accuracy value is 83.33%. For more details can be seen in Figure 12

```

- SUPPORT VECTOR MACHINE CLASSIFIER (SVM)

[63] # IMPORT LIBRARY ALGORITMA
    from sklearn import svm

[64] # INISIASI MODEL
    svmc = svm.SVC()

# MELAKUKAN LATIH/ FITTING MODEL
svmc.fit(X_train, Y_train)

SVC()

[65] # MELAKUKAN TESTING/UJI MODEL
    y_pred_svm = svmc.predict(X_test)

[66] # MELIHAT HASIL/NILAI UJI/TESTING MODEL
    accuracy_svmc = accuracy_score(Y_test, y_pred_svm)

# Jika Nilai Testing > Nilai Training maka disebut Overfitting
print("Nilai Akurasi Model SVM dalam Dataset Training adalah: {:.3f}".format(svmc.score(X_train, Y_train)*100))
print("Nilai Akurasi Model SVM dalam Dataset Testing adalah: ", accuracy_svmc*100)

Nilai Akurasi Model SVM dalam Dataset Training adalah: 88.703
Nilai Akurasi Model SVM dalam Dataset Testing adalah: 83.33333333333334
    
```

Figure 12 Training and testing models

To get a comparison of accuracy, hyper tuning is performed on the Support Vector Machine (SVM) with parameter "C" : [0.1, 1, 10, 100, 1000], "gamma" : [1, 0.1, 0.01, 0.001, 0.00001], "kernel" : ['rbf']. Where the best parameter generated after tuning is "C": 100, "gamma": 0.001, "kernel": 'rbf'. Then the accuracy results obtained for testing after tuning are 81.66%, as shown in Figure 13.

```

0s ✓ # DEKLARASI PARAMETER TUNING UNTUK DIBANDINGKAN
    param_grid = {"C" : [0.1, 1, 10, 100],
                  "gamma" : [1, 0.1, 0.01, 0.001],
                  "kernel" : ['rbf']}

# INISIALISASI HYPERTUNING
grid_svm = GridSearchCV(svm.SVC(), param_grid, refit=True, verbose=3)

# FITTING MODEL SVM DENGAN HYPERTUNING
grid_svm.fit(X_train, Y_train)

[69] # MELIHAT PARAMETER / PART TUNING TERBAIK UNTUK SVM

print("Parameter Terbaik Setelah Tuning :\n", grid_svm.best_params_)

Parameter Terbaik Setelah Tuning :
{'C': 100, 'gamma': 0.001, 'kernel': 'rbf'}
GridSearchCV: grid_svm
sklearn.model_selection._search.GridSearchCV instance

# MELAKUKAN TESTIN
y_pred_grid_svm = grid_svm.predict(X_test)

[71] # MELIHAT AKURASI MODEL
accuracy_grid_svm = accuracy_score(Y_test, y_pred_grid_svm) * 100
accuracy_grid_svm

81.66666666666667
    
```

Figure 13 Optimization/hyper tuning model

3.7 Model Evaluation

The results of the Support Vector Machine (SVM) algorithm evaluation model before tuning are that the model can guess class 0 correctly 40 times out of 5 attempts, and the model can guess class 1 as much as 5 times correct out of 10 trials. For visualization of the Support Vector Machine confusion matrix before tuning, can be seen in Figure 14.

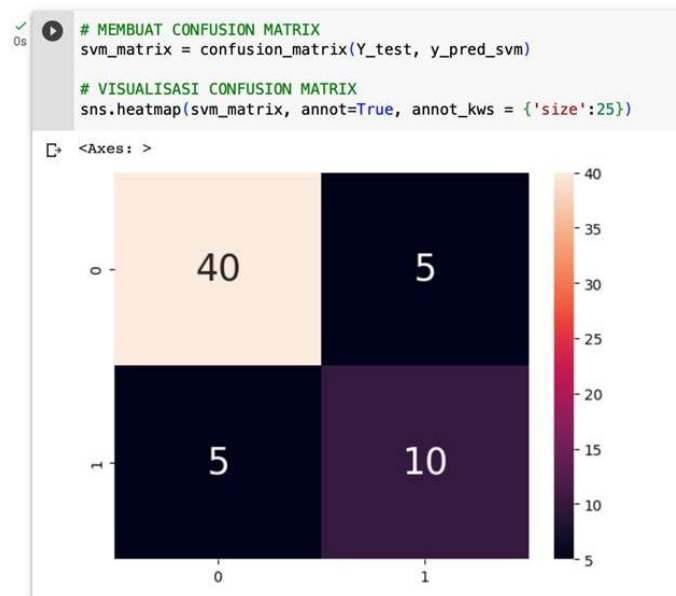


Figure 14 Confusion matrix Support Vector Machine (SVM)

While the results of the Support Vector Machine (SVM) algorithm evaluation model after tuning are that the model can guess class 0 correctly 38 times out of 7 attempts and can

guess class 1 precisely 4 times out of 11 trials. For visualization of the Support Vector Machine confusion matrix after tuning, can be seen in Figure 15.

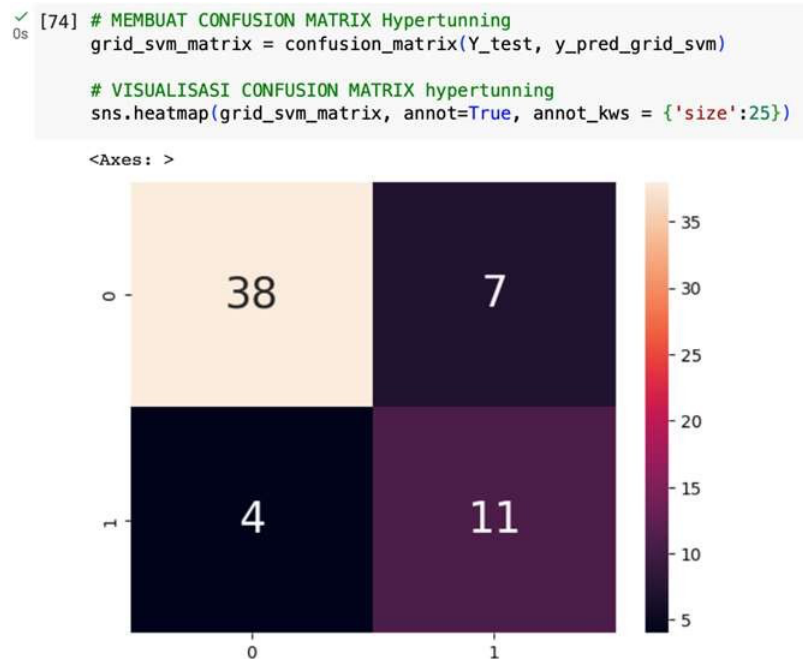


Figure 15 Confusion matrix Support Vector Machine (SVM) after tuning

3.8 Interpretation of Results

The classification results before and after tuning can be seen in Figure 16. There is a slight difference in the accuracy value of the f1-score before tuning is better. The accuracy value of the f1-score before tuning is 0.83, while after adjusting it is 0.82.

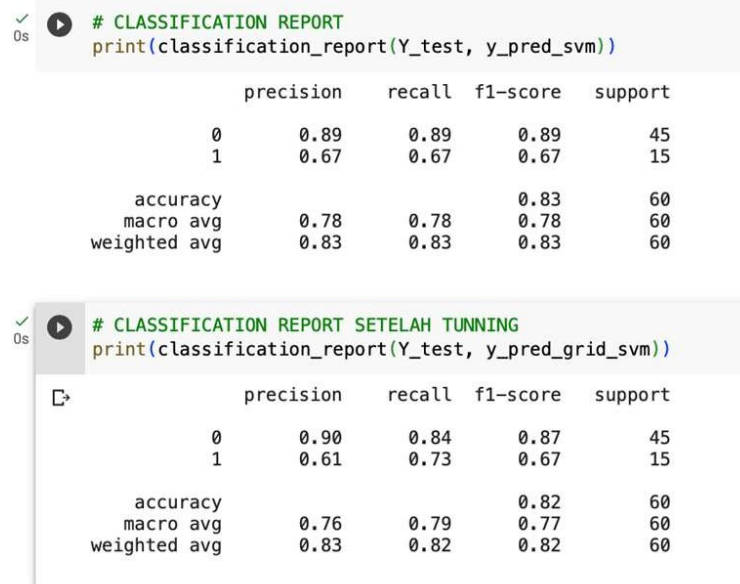


Figure 16 Classification Results

CONCLUSION

This study shows that the Support Vector Machine (SVM) has a high degree of accuracy in predicting heart failure risk classification. The results of this study provide confidence that the classification model built can be an effective tool in assisting the diagnosis and management of

heart failure patients.

In conclusion, this classification model has good performance with an F1-Score accuracy of 0.83 in predicting the risk of heart failure based on clinical data. This can help medical personnel identify high-risk patients and provide appropriate treatment or intervention to prevent disease progression and improve prognosis.

In the future, similar research can be carried out by increasing the dataset and using other methods as comparisons to obtain more information about heart failure. That way, it can help medical personnel in treating heart failure patients so that the death rate from heart failure can be reduced.

BIBLIOGRAPHY

- [1] Fredilio, J. Rahmad, S. Hamonangan Sinurat, D. Ryan Hamonangan Sitompul, D. Jusuf Ziegel, and E. Indra, "Comparison of the K-Nearest Neighbors (K-NN) Algorithm and Random Forest for Heart Failure," *Journal of Technology Informatics and Computers MH. Thamrin*, vol. 9, no. 1, pp. 471–486, Mar. 2023, doi: 10.37012/jtik.v9i1.1432.
- [2] APLumi, VFF Joseph, and NCI Polii, "Cardiac Rehabilitation in Patients with Chronic Heart Failure," *Journal of Biomedicine:JBM*, vol. 13, no. 3, p. 309, Apr. 2021, doi: 10.35790/jbm.v13i3.33448.
- [3] R.Indrakumari, T. Poongodi, and SR Jena, "Heart Disease Prediction using Exploratory Data Analysis," in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.06.017.
- [4] S.Aldera, A. Emam, M. Al-Qurishi, M. Alrubaiyan, and A. Alothaim, "Exploratory Data Analysis and Classification of a New Arabic Online Extremism Dataset," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3132651.
- [5] K. Sahoo*, AKSamal, J. Pramanik, and SK Pani, "Exploratory Data Analysis using Python," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, p. 4727–4735, Oct. 2019, doi: 10.35940/ijitee.L3591.1081219.
- [6] A. Kulkarni and A. Shivananda, *Natural Language Processing Recipes*. Berkeley, CA: Apress, 2019. doi: 10.1007/978-1-4842-4267-4.
- [7] A. Arifin, J. Hendyli, and DE Herwindiati, "Classification of Herbal Medicinal Plants Using the Support Vector Machine Method," *Computatio : Journal of Computer Science and Information Systems*, vol. 5, no. 1, 2021, doi: 10.24912/computatio.v1i1.12811.
- [8] QHasanah, H. Oktavianto, and YD Rahayu, "The Gaussian Naive Bayes Algorithm Analysis of Data Classification of Heart Failure Patients," *Journal of Smart Technology*, vol. 3, no. 4, pp. 382–389, 2022.
- [9] S. Saida, H. Haryati, and L. Rangki, "Quality of Life of Patients with Congestive Heart Failure Based on Degree of Physical Ability and Duration of Disease," *Faletehan Health Journal*, vol. 7, no. 02, pp. 70–76, Jul. 2020, doi: 10.33746/fhj.v7i02.134.
- [10] IP Princess, "Performance Analysis of the K-Nearest Neighbor (KNN) Method and Crossvalidation in Cardiovascular Disease Data," *Indonesian Journal of Data and Science*, vol. 2, no. 1, pp. 21–28, Mar. 2021, doi: 10.33096/ijodas.v2i1.25.
- [11] H. Christanto, J. Rahmad, S. Hamonangan Sinurat, D. Ryan Hamonangan Sitompul, A. Sitomorang, and D. Jusuf Ziegel, "Comparative Analysis of Decision Trees, Support Vector Machines, and Xgboost in Classifying Trip Advisor Hotel Reviews," *Journal of Informatics Technology and MH Computers. Thamrin*, vol. 9, no. 1, pp. 306–319, 2023, doi: 10.37012/jtik.v9i1.1429.
- [12] R. Kotha, "Heart Failure Clinical Records Classification," Kaggle, 2021.
- [13] DTHusni et al., "BIG DATA SALES ANALYSIS VIDEOS GAMES USING EDA," *Journal of Information and Computer Engineering (Tekinkom)*, vol. 5, no. 1, p. 43, Jun. 2022, doi: 10.37600/tekinkom.v5i1.517.
- [14] DYSiringo, V. Sihombing, and M. Masrizal, "SYSTEM SALES AND INVENTORY

INFORMATIONWEB-BASED AGRICULTURAL EQUIPMENT PRODUCTS,” Journal of Computer and Information Engineering (Tekinkom), vol. 4, no. 1, pp. 54–59, 2021,doi: 10.37600/tekinkom.v4i1.232.

- [15] M. Radhi, D. RyanHamonangan Sitompul, S. Hamonangan Sinurat, and E. Indra, "PREDICTION OF CAR PRICE USING REGRESSION ALGORITHMS WITH HYPER-PARAMETER TUNING," Journal of Information Systems and Computer Science Prima, vol. 4, no. 2, 2021.
- [16] MM Hassan, MZ Uddin, A. Mohamed, andA Almogren, “A robust human activity recognition system using smartphone sensors and Deep Learning,” Future Generation Computer Systems, vol. 81, pp. 307–313, 2018.
- [17] A. Kumar, “Learning predictive analytics with python: Gain practical insights into predictive modeling by implementing predictive analytics algorithms on public datasets withpython,” Packt Publishing, 2016.
- [18] R.Officialati and T. Arifin, "Classification of Breast Cancer Patients Using the Support Vector Machine Method with Backward Elimination," SYSTEMASI, vol. 10, no. 2, 2021, doi: 10.32520/stmsi.v10i2.1238.
- [19] Oryza Habibie Rahman,Gunawan Abdillah, and Agus Komarudin, "Classification of Hate Speech on Twitter Social Media Using Support Vector Machine," RESTI Journal (System Engineering and Information Technology), vol. 5, no. 1, 2021, doi: 10.29207/resti.v5i1.2700.
- [20] A. Y and KY S, “Multiclass Classification and Support Vector Machine,” Global Journal of Computer Science and Technology Interdisciplinary, vol. 12, no. 11, p. 14–19, 2012.
- [21] R.Moraes, JF Valiati, and WP Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Expert System Appl, vol. 40., no. 2, pp. 621–633, 2013.