VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i2.5719

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

UNLEASHING THE POWER OF SVM AND KNN: ENHANCED EARLY DETECTION OF HEART DISEASE

Jefri Junifer Pangaribuan[1*]; Ade Maulana[2]; Romindo[3]

Information Systems[1,2,3]
Universitas Pelita Harapan, Medan, Indonesia[1,2,3]
https://www.uph.edu/en/[1,2,3]
jefri.pangaribuan@uph.edu[1*], ade.maulana@lecturer.uph.edu[2], romindo@uph.edu[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Heart disease is a fatal illness responsible for approximately 36% of deaths in 2020. Therefore, it is important to pay attention to and better anticipate the risk of heart disease. One technological contribution that can be made is through information related to the risk of heart disease. Classification techniques in data mining can be used to diagnose and identify the risk of heart disease earlier by processing medical data and making predictions. This study compares the effectiveness of two classification algorithms, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), in predicting the risk of heart disease using a Kaggle dataset consisting of 303 records with 14 attribute columns. The data is divided into 70% for training and 30% for testing. The software used in this study is Orange Data Mining to build the SVM and KNN models. The results show that the SVM accuracy is 85.6%, while KNN achieves 81.1%. Based on the confusion matrix, the SVM algorithm has a lower error rate compared to KNN. In conclusion, the SVM algorithm is superior to KNN in predicting the risk of heart disease. These findings indicate that SVM has a better potential in identifying individuals at high risk of experiencing a heart attack. This research can contribute to the development of a more accurate medical decision support system for early detection of heart disease.*

Keywords: *confusion matrix, heart attack, k-nearest neighbor, support vector machine.*

Intisari— *Serangan jantung adalah penyakit mematikan yang menyebabkan sekitar 36% kematian pada tahun 2020. Oleh karena itu, penting untuk memperhatikan risiko serangan jantung dan mengantisipasinya dengan lebih baik. Salah satu kontribusi teknologi yang dapat diberikan adalah melalui informasi terkait risiko serangan jantung. Teknik klasifikasi dalam data mining dapat digunakan untuk mendiagnosis dan mengidentifikasi risiko serangan jantung lebih awal dengan memproses data medis dan membuat prediksi. Penelitian ini membandingkan efektivitas dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN), dalam memprediksi risiko serangan jantung menggunakan dataset dari Kaggle yang terdiri dari 303 data dengan 14 kolom atribut. Data tersebut dibagi menjadi 70% untuk data training dan 30% untuk data testing. Software yang digunakan dalam penelitian ini adalah Orange Data Mining untuk membangun model SVM dan KNN. Hasil penelitian menunjukkan bahwa akurasi SVM mencapai 85,6%, sementara KNN mencapai 81,1%. Berdasarkan confusion matrix, algoritma SVM menghasilkan tingkat kesalahan yang lebih kecil dibandingkan KNN. Hasil ini mengindikasikan bahwa SVM memiliki potensi yang lebih baik dalam mengidentifikasi seseorang yang berisiko tinggi mengalami serangan jantung. Temuan ini dapat berkontribusi pada pengembangan sistem pendukung keputusan medis yang lebih akurat untuk deteksi dini penyakit jantung.*

Kata Kunci: *confusion matrix, serangan jantung, k-nearest neighbor, support vector machine.*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5719

INTRODUCTION

One of the deadliest diseases in the world is a heart attack, which occurs due to a blockage in the coronary artery blood flow, causing the oxygen level in the heart muscle to decrease until an infarction occurs [1]. Factors that contribute to heart attacks include smoking habits, poor sleep quality, and the consumption of high-cholesterol foods, which have become common among people as daily lifestyles change [2]. Research has shown that in Indonesia in 2012, there were approximately 43.30% of people affected by heart attacks, leading to a mortality rate of 12.90% [3]. In 2013, it was recorded that 61,682 people were affected by heart attacks, with the majority of cases occurring in Central Java, totaling 11,511 cases. This data is further supported by research findings which estimate that by 2020, heart disease would become the leading cause of death worldwide, accounting for 36% of all deaths. This figure is considered to be twice as high as the mortality rate from cancer [4].

This condition can occur at any time, which is why immediate treatment is necessary to prevent death. Common symptoms of a heart attack are [5]:

1. Irregular heartbeat
2. Shortness of breath
3. Chest feels like it's being pressed by a heavy object
4. Pain in the chest that radiates to the back and arms
5. Nausea, vomiting, and weakness
6. Excessive cold sweating

Other factors that contribute to heart attacks are as follows [6]:

1. Age (Men >45 years and Women >55 years)
2. High blood pressure (hypertension)
3. Cholesterol
4. High blood sugar
5. Stress
6. Use of illegal drugs

The data above proves that heart disease is very dangerous. Therefore, this research uses heart disease patient data obtained from the website Kaggle.com. Heart disease data is too extensive for manual processing, so it will be processed through data mining using machine learning [7]. Data mining techniques are related to fields of science and technology such as data statistics, artificial intelligence, and machine learning [8]. In data mining, there are several data processing techniques used, one of which is classification. Researchers often use classification as a process to place certain objects into a grouping divided into several categories [9]. In the current research, the classification method is used to predict whether a person will have a heart attack so that early treatment can be administered. Based on the research conducted, the classification of heart disease using the K-Nearest Neighbor (KNN) algorithm yielded an accuracy rate of 88.333% in the first trial and 96.667% in the second trial [10]. According to the research conducted, the classification of heart disease using the Support Vector Machine (SVM) algorithm yielded an accuracy rate of 87% in the first trial and 80% in the second trial [11].

In the management of classification data, this research uses algorithms such as SVM and KNN. The SVM algorithm is a classification method for both linear and non-linear data that systematically utilizes the function of a kernel to transform the support vector classifier into a higher-dimensional form [12]. The working principle of this method is to determine the maximum distance from a hyperplane by iteratively adjusting until the optimal hyperplane distance is achieved for both classes. The KNN algorithm is a classification method used by collecting several objects from the data that have undergone the training process, which have the closest distance and highest similarity to the test data [13]. Using these two algorithms, the comparison reveals that both SVM and KNN have the same advantages: simple and commonly used data grouping processes, ease of implementation, and good performance in their methods [14].

Furthermore, SVM and KNN have been employed in various studies, either by comparing the two methods or combining them. Some of these studies include comparing air quality in Jakarta [15], predicting student’s performance [16], classifying lung lesion scheme [17], recognizing speech emotions [18], and classifying brain tumors [19].

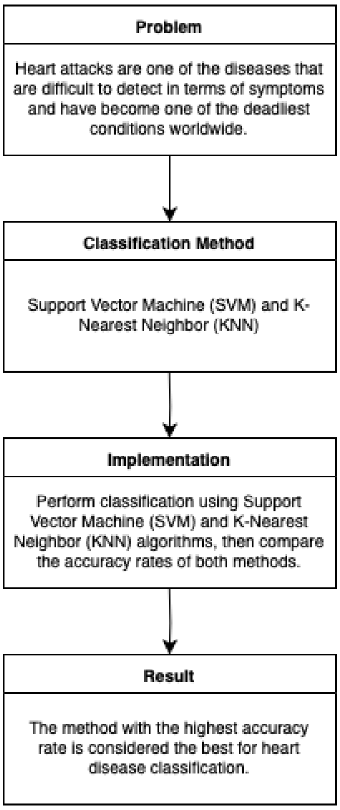
Therefore, this research aims to evaluate and compare the effectiveness of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms in classifying heart disease using datasets obtained from Kaggle.com. By implementing these classification algorithms, the study seeks to determine which algorithm provides superior accuracy in detecting heart disease. The findings from this comparative analysis will contribute to optimizing predictive models for early detection and management of heart disease, ultimately aiding in improving health outcomes and guiding future research in the field.

MATERIALS AND METHODS

To achieve the research objectives already explained, a conceptual framework was created.

This study discusses the comparative analysis of the SVM and KNN algorithms in detecting heart disease, focusing on these two algorithms. The data used comes from the Kaggle website and is then classified using the SVM and KNN algorithms with the help of the Orange Data Mining application. The results obtained are the accuracy rates of the SVM and KNN algorithms, and these accuracy rates are compared to determine which algorithm is the most accurate in detecting heart disease.

The conceptual framework depicted in the diagram below is as follows:



Source: (Research Results, 2024)
Figure 1. Conceptual Framework

A. Data Collection Methods

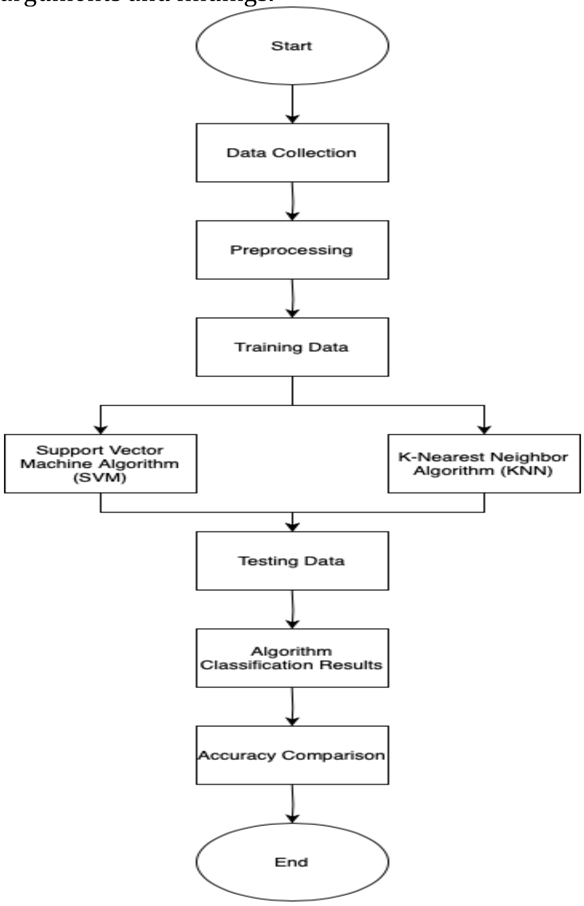
In this study, the data collection method used is through internet searching. This data collection method utilizes internet search engine technology to gather information and data. The data collected originates from the Kaggle website, specifically the "Heart Attack Analysis & Prediction Datasets," which provides analysis and predictions of patients exhibiting symptoms of heart disease [20].

B. Hypothesis

In this study, two of several machine learning algorithm methods for classification will be used to detect heart disease, namely SVM and KNN. The author hypothesizes that the accuracy rate provided by the SVM algorithm will be greater than that of the KNN algorithm.

C. Solution Methods

The stages of the research are conducted to support the flow of the study and provide strong arguments and findings.



Source: (Research Results, 2024)
Figure 2. Research Stages

Figure 2 shows the stages of the research, starting with data collection through searching for datasets to be used for classification. The data obtained originates from the Kaggle website, containing various criteria data of heart attack patients. The next stage involves data pre-processing, where data cleaning ensures no duplicate data, fills in missing values, and smooths noisy data. Subsequently, the data is normalized to ensure no excessive data and is generalized, transforming the data into a suitable format for the mining process.

The data is divided for training and testing purposes. A percentage of 70% of the data is used for training, and 30% for testing. The next stage involves the classification of SVM and KNN algorithms, using various model calculations of these algorithms on the training data.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5719

The results of the SVM and KNN algorithm classifications provide accuracy rates that will be compared to determine the best accuracy rate in detecting heart disease.

D. Prediction of SVM Algorithm

The SVM algorithm is a commonly used machine learning method for analyzing and categorizing data into two categories. In this research, the SVM algorithm uses training and testing data to perform classification in order to predict heart disease.

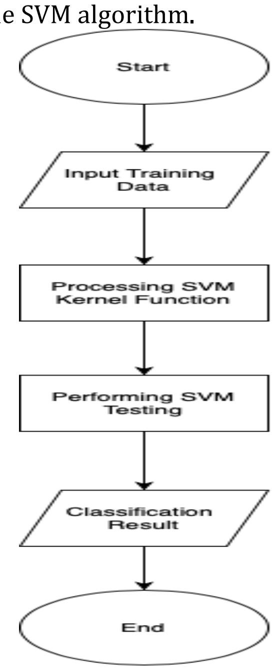
The steps of the SVM algorithm base on figure 2 are as follows:

1. Inputting training data for training purposes. The training process involves using separated data, then training using the SVM algorithm.
2. The application process of the SVM kernel function involves using the Polynomial kernel. The equation for the Polynomial kernel function is as follows:

$$K(x_1, x_2) = (x_1 \cdot x_2 + 1)^d \quad (1)$$

x_1 = trained data
x_2 = test data
d = polynomial degree parameter (typically d=2)

3. After training the data, testing is conducted to evaluate the SVM classification algorithm and obtain accuracy values. Evaluate, analyze, and discuss the results of the classification when using the SVM algorithm.

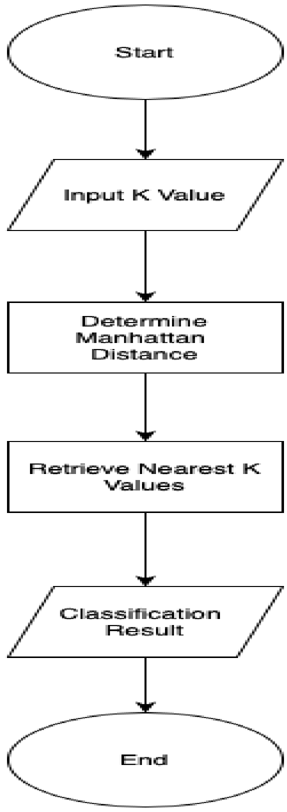


Source: (Research Results, 2024)
Figure 3. SVM Flowchart

E. Prediction of KNN Algorithm

The KNN algorithm is a data mining classification method and is also part of supervised learning. The KNN algorithm classifies objects based on attributes derived from query instances categorized by the majority of a category in that algorithm. In this study, the KNN algorithm is built using training and testing data to classify data for predicting heart attack patients.

The operational steps of the KNN algorithm are as follows.



Source: (Research Results, 2024)
Figure 4. KNN Flowchart

In the picture above is the flowchart of the KNN algorithm. The steps of how the KNN algorithm works are as follows:

1. Enter the minimum value of k, k = 1, and the maximum value which can be given as the number of trained data.
2. Determine the Manhattan distance. The calculation in finding the Manhattan distance is as follows:

$$d_{Manhattan}(x, y) = \sum_{i=1}^{n} |x_1 - y_1| \quad (2)$$

i = number of data sets
x = number of test data
y = number of trained data

3. Retrieve the nearest k values according to the calculated distance and calculate the data points in each category.
4. Assign the new data point to a category with the highest number of neighbors.
5. Determine the classification results from the fourth step with the highest number.

F. Confusion Matrix

The Confusion Matrix is a performance measurement tool in machine learning classification problems that generates 4 different combinations of predicted and actual values. These four terms represent True Positive, True Negative, False Positive, and False Negative outcomes.

Table 1. Confusion Matrix

		Predicted	
		Negative (-)	Positive (+)
Actual	Negative (-)	True Negative (TN)	False Positive (FP)
	Positive (+)	False Negative (FN)	True Positive (TP)

Source: (Research Results, 2024)

1. True Positive (TP): predicted positive value that is correct.
2. True Negative (TN): predicted negative value that is correct.
3. False Positive (FP): predicted positive value that is incorrect.
4. False Negative (FN): predicted negative value that is incorrect.

In the Confusion Matrix, it includes Accuracy, Recall, Precision, and F-Score explained below [21]:

1. Accuracy
   A metric used to measure the classification model that provides all correct predictions.
   $$AC = \frac{TP+TN}{\mathrm{TP+TN+FP+FN}} x100\% \quad (3)$$
   AC: Accuracy
   TP: True Positive
   TN: True Negative
   FP: False Positive
   FN: False Negative

2. Recall
   The metric used to measure how many positive instances are identified by the model.
   $$RC = \frac{TP}{\mathrm{TP+FN}} x100\% \quad (4)$$
   RC: Recall
   TP: True Positive
   FN: False Negative

3. Precision
   The metric used to measure how many positive predictions obtained from the model are correctly stated.
   $$PR = \frac{TP}{\mathrm{TP+FP}} x100\% \quad (5)$$
   PR: Precision
   TP: True Positive
   FP: False Positive

4. F1-Score
   A combined metric that considers both precision and recall, aiming to achieve a balance between these two metrics.
   $$F1 = 2x \frac{RC\ x\ PR}{\mathrm{RC+PR}} x100\% \quad (6)$$
   RC: Recall
   PR: Precision

RESULTS AND DISCUSSION

A. Data Collection Results

After collecting data from the Kaggle website, a dataset of heart attack patients with classification properties was obtained. This dataset contains 303 patients and is provided in a Comma Separated Values (CSV) file. The dataset includes attribute columns and information related to heart attack patients. The data has 14 attributes, which can be seen in the table below.

Table 2. Heart Attack Data Attributes

Attribute Name	Attribute Description
Age	Age of the patient
Sex	Gender of the patient (example = 0 for female, 1 for male).
Chest Pain	Type of chest pain or chest pain symptoms experienced by the patient. (0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic).
Resting Blood Pressure	Blood pressure at rest
Cholesterol	Cholesterol level in the blood
Fasting Blood Sugar	Blood sugar level (example = 0 for normal blood sugar, 1 for high blood sugar).
Resting Electrocardiographic	Resting electrocardiogram (EKG or ECG) results (0 = normal, 1 = ST wave deviation greater than 0.5mV, 2 = hypertrophy)
Maximum Heart Rate Achieved	Maximum heart rate achieved during exercise test
Exercise Induced Angina	Whether the patient experiences angina (chest pain) induced by exercise (example = 0 for no, 1 for yes)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5719

Attribute Name	Attribute Description
ST Depression Induced	ST segment depression on the electrocardiogram induced by exercise
Slope	Slope of the ST segment on the electrocardiogram during exercise or exercise test
Number of Major Vessels	Number of major vessels colored by contrast agent during cardiac angiography
Thalassemia	Blood disorder (example = 0 for normal, 1,2, or 3 for different types)
Output	Whether the patient has a chance of having a heart attack. If "0," there is a low chance of a heart attack, whereas "1" indicates a high chance of a heart attack.

Source: (Research Results, 2024)

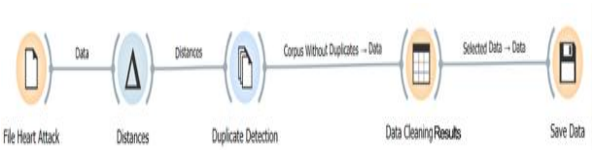
Some attributes in the table above refer to factors and causes of heart disease. These attributes are often used in related research to identify risk factors and aid in predicting and diagnosing heart disease.

B. Research Results

To train the SVM and KNN algorithms with the Orange Data Mining application, the heart attack patient dataset needs to undergo data cleaning and transformation first to ensure stable performance during classification.

1. Data Cleaning

This dataset has 303 entries with 14 attributes columns.



Source: (Research Results, 2024)

Figure 5. Data Cleaning Setup

In Figure 5, the data will be processed to remove duplicate entries. The initial step involves inputting the heart attack patient dataset into a file to display the attributes and the number of entries. Then, the distance feature is used to calculate the distance between columns and rows of the dataset. Next, the process moves to duplicate detection, where duplicate data will be cleaned, and the data is displayed in a table format to review all the data after cleaning.

2. Data Transformation

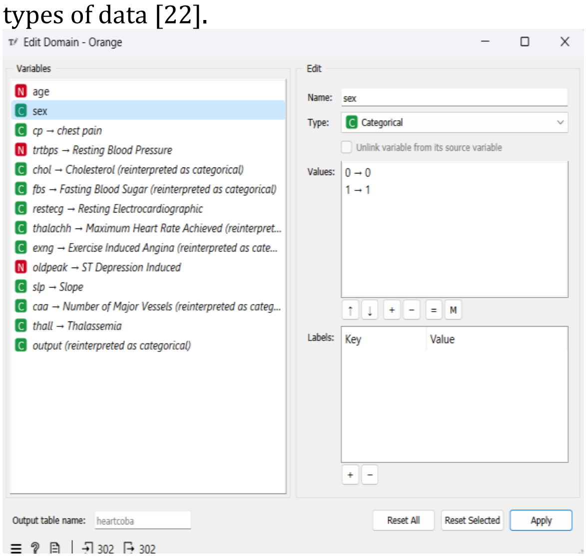
In Table 3, it shows that the initial data attributes are still abbreviated, with types being numeric and categorical, serving as features, and some attributes having values of 0 and 1.

Table 3. Initial Data Attributes

	Name	Type	Role	Values
1	age	numeric	feature	
2	sex	categorical	feature	0, 1
3	cp	categorical	feature	
4	trtbps	numeric	feature	
5	chol	numeric	feature	
6	fbs	numeric	feature	0, 1
7	restecg	categorical	feature	
8	thalachh	numeric	feature	
9	exng	numeric	feature	0, 1
10	oldpeak	numeric	feature	
11	slp	categorical	feature	
12	caa	numeric	feature	
13	thall	categorical	feature	
14	output	numeric	feature	0, 1

Source: (Research Results, 2024)

Then in Figure 6, changes in several attribute values and names are shown. Attributes such as Age, Trtbps (Resting Blood Pressure), and Oldpeak (ST Depression Induced) remain unchanged because modifying them would degrade the performance of the predictive algorithm to be used. The method used to categorize data categories employs Label Encoding, where categories are assigned values starting from 0 onwards to categorize various types of data [22].



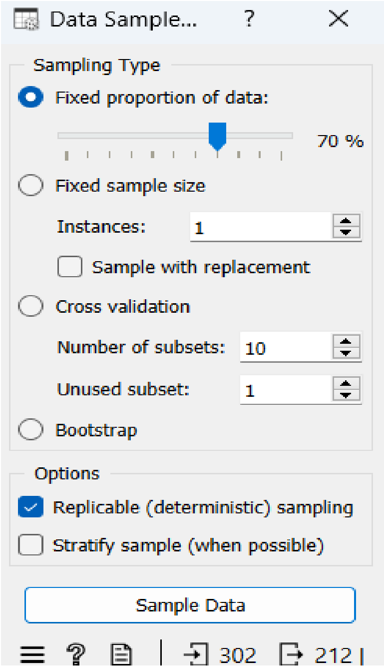
Source: (Research Results, 2024)

Figure 6. Changes in Initial Data Attributes

3. Data Sampler

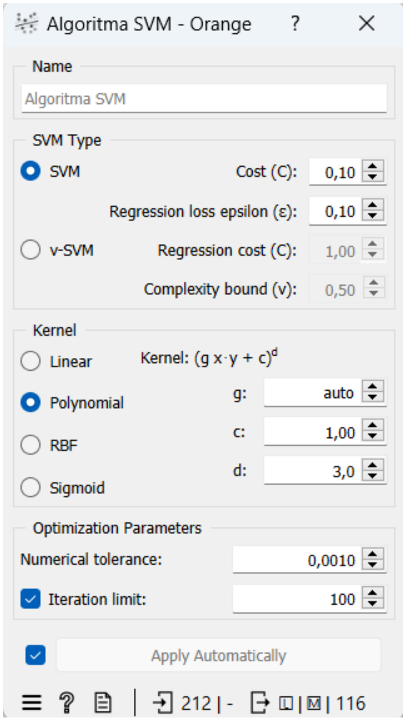
The data sampler is a component used in data processing to sample data to understand and analyze its characteristics. The data sampler uses the fixed proportion of data technique to take 70% of the data as training data, with the remainder used as testing data.

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i2.5719
JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)



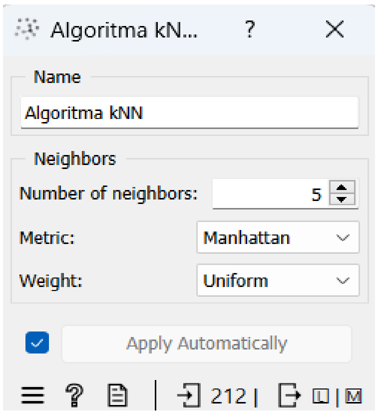
Source: (Research Results, 2024)
Figure 7. Changes in Initial Data Attributes

4. Building SVM and KNN Algorithms
The SVM algorithm uses the SVM type with a Polynomial kernel. The kernel uses g (gamma) =auto, c0 = 1 (default), and d (degree) = 3 (default).



Source: (Research Results, 2024)
Figure 8. SVM Algorithm Setup

The KNN algorithm uses 5 neighbors, with Manhattan distance, and uniform weights.



Source: (Research Results, 2024)
Figure 9. KNN Algorithm Setup

5. Prediction Results of SVM and KNN Algorithms
After classification using the remaining 30% of the data from the training data, the prediction results can be seen in the image below. In Figure 10, it shows that the SVM algorithm has a superior accuracy of 85.6% compared to KNN, which has an accuracy of 81.1%.

	Algoritma kNN	Algoritma SVM	output
1	0.80 : 0.20 → Less Chance	0.83 : 0.17 → Less Chance	More Chance
2	0.20 : 0.80 → More Cha...	0.05 : 0.95 → More Cha...	More Chance
3	0.00 : 1.00 → More Cha...	0.04 : 0.96 → More Cha...	More Chance
4	0.00 : 1.00 → More Cha...	0.06 : 0.94 → More Cha...	More Chance
5	0.20 : 0.80 → More Cha...	0.78 : 0.22 → Less Chance	Less Chance
6	0.80 : 0.20 → Less Chance	0.04 : 0.96 → More Cha...	More Chance
7	0.20 : 0.80 → More Cha...	0.50 : 0.50 → More Cha...	More Chance
8	0.60 : 0.40 → Less Chance	0.96 : 0.04 → Less Chance	Less Chance
9	0.40 : 0.60 → More Cha...	0.05 : 0.95 → More Cha...	More Chance
10	0.00 : 1.00 → More Cha...	0.07 : 0.93 → More Cha...	More Chance
11	0.60 : 0.40 → Less Chance	0.96 : 0.04 → Less Chance	Less Chance
12	0.00 : 1.00 → More Cha...	0.06 : 0.94 → More Cha...	More Chance
13	0.20 : 0.80 → More Cha...	0.10 : 0.90 → More Cha...	More Chance
14	0.40 : 0.60 → More Cha...	0.02 : 0.98 → More Cha...	More Chance
15	0.80 : 0.20 → Less Chance	0.96 : 0.04 → Less Chance	Less Chance
16	1.00 : 0.00 → Less Chance	0.97 : 0.03 → Less Chance	Less Chance
17	0.40 : 0.60 → More Cha...	0.22 : 0.78 → More Cha...	More Chance
18	0.40 : 0.60 → More Cha...	0.42 : 0.58 → More Cha...	More Chance
19	0.60 : 0.40 → Less Chance	0.82 : 0.18 → Less Chance	Less Chance
20	0.60 : 0.40 → Less Chance	0.35 : 0.65 → More Cha...	More Chance
21	0.00 : 1.00 → More Cha...	0.05 : 0.95 → More Cha...	More Chance
22	0.60 : 0.40 → Less Chance	0.87 : 0.13 → Less Chance	Less Chance
23	0.60 : 0.40 → Less Chance	0.91 : 0.09 → Less Chance	Less Chance
24	0.20 : 0.80 → More Cha...	0.05 : 0.95 → More Cha...	More Chance
25	0.40 : 0.60 → More Cha...	0.87 : 0.13 → Less Chance	Less Chance
26	0.60 : 0.40 → Less Chance	0.98 : 0.02 → Less Chance	Less Chance
27	1.00 : 0.00 → Less Chance	0.99 : 0.01 → Less Chance	Less Chance

Show perfomance scores   Target class: (Average over classes)

Model	AUC	CA	F1	Precision	Recall
Algoritma kNN	0.853	0.811	0.811	0.812	0.811
Algoritma SVM	0.916	0.856	0.855	0.856	0.856

Source: (Research Results, 2024)
Figure 10. Prediction Results of SVM and KNN

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5719

C. Discussion

When predicting heart attacks on the testing data using the SVM algorithm, it provided an Area under ROC Curve of 91.6%, an Accuracy of 85.6%, an F1-Score of 85.5%, a Precision of 85.6%, and a Recall of 85.6%. In comparison, the KNN algorithm provided an Area under ROC Curve of 85.3%, an Accuracy of 81.1%, an F1-Score of 81.1%, a Precision of 81.2%, and a Recall of 81.1%. It can be stated that the SVM algorithm is superior in predicting heart disease with higher accuracy compared to the KNN algorithm.

The Confusion Matrix produced in the study using the SVM and KNN algorithms on the Testing Data.

Actual \ Predicted	Less Chance	More Chance	Σ
Less Chance	31	8	39
More Chance	5	46	51
Σ	36	54	90

Source: (Research Results, 2024)

Figure 11. SVM Confusion Matrix

In Figure 11, it shows that the SVM algorithm produces the following prediction values: 46 true positives (correctly predicted as low risk of heart attack), 8 false positives (predicted as high risk but actually low risk), 31 true negatives (correctly predicted as high risk), and 5 false negatives (predicted as low risk but actually high risk).

Actual \ Predicted	Less Chance	More Chance	Σ
Less Chance	31	8	39
More Chance	9	42	51
Σ	40	50	90

Source: (Research Results, 2024)

Figure 12. KNN Confusion Matrix

In Figure 12, it shows that the KNN algorithm produces the following prediction values: 42 true positives (correctly predicted as low risk of heart attack), 8 false positives (predicted as high risk but actually low risk), 31 true negatives (correctly predicted as high risk), and 9 false negatives (predicted as low risk but actually high risk).

From the Confusion Matrix results described above, it can be concluded that the SVM algorithm's predictions are superior to the KNN algorithm, as the SVM has lower prediction errors compared to the KNN.

CONCLUSION

In conclusion, the research findings indicate that the Support Vector Machine (SVM) algorithm achieves an accuracy of 85.6%, while K-Nearest Neighbor (KNN) records an accuracy of 81.1% in predicting heart disease. This demonstrates that SVM performs better than KNN in this particular dataset. These results provide valuable insights into the use of machine learning for heart disease prediction, suggesting that SVM may be a more reliable option for this task. Further exploration into other datasets and model optimization could enhance the accuracy and robustness of these predictions in real-world medical applications.

Future research should consider testing other machine learning algorithms to explore their effectiveness in heart disease detection. Increasing the size and diversity of the dataset could also improve prediction accuracy and generalizability. Additionally, applying techniques like hyperparameter optimization, feature selection, and ensemble methods could further refine model performance, ensuring more accurate and reliable outcomes in medical diagnoses.

ACKNOWLEDGEMENT

This article is based on research funded by the Research and Community Service Institute (Lembaga Penelitian dan Pengabdian kepada Masyarakat) of Pelita Harapan University.

REFERENCE

[1] D. Kurniawan, K. Ibrahim, and A. Prawesti, "Pengalaman Pasien Mengalami Serangan Jantung Pertama Kali yang Dirawat di ruang CICU," *Jurnal Keperawatan Padjadjaran*, vol. v3, no. n2, pp. 67–76, Aug. 2015, doi: 10.24198/jkp.v3n2.2.

[2] F. Saktiningtyastuti, L. Dwi, A. Kementerian, K. Politeknik, K. Surakarta, and J. Keperawatan, "Faktor yang Mempengaruhi Serangan Jantung Berulang pada Pasien AMI di Ruang ICVCU RSUD Dr. Moewardi Tahun 2016," *(JKG) JURNAL KEPERAWATAN GLOBAL*, vol. 2, no. 1, Jun. 2017, Accessed: Jul. 04, 2024. [Online]. Available: http://jurnal.poltekkes-solo.ac.id/index.php/JKG/article/view/346

[3] M. A. Bianto, K. Kusrini, and S. Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes," *Creative Information Technology Journal*, vol. 6, no. 1,

p. 75, Apr. 2020, doi: 10.24076/citec.2019v6i1.231.

[4] L. Baransyah, M. S. Rohman, and T. Suharsono, "Faktor-Faktor yang Berpengaruh Terhadap Kejadian Gagal Jantung pada Pasien Infark Miokard Akut di Rumah Sakit Dr. Saiful Anwar Malang," *Majalah Kesehatan* , vol. 1, no. 4, pp. 209–213, 2014, Accessed: Jul. 04, 2024. [Online]. Available: https://majalahfk.ub.ac.id/index.php/mkfkub/article/view/42

[5] S. Hospitals, "Pertolongan Pertama Serangan Jantung yang Wajib Diketahui." Accessed: Jul. 04, 2024. [Online]. Available: https://www.siloamhospitals.com/informasi-siloam/artikel/pertolongan-pertama-serangan-jantung

[6] Dr. F. Rizal Makarim, "Apa itu Serangan Jantung? Gejala, Penyebab, dan Pengobatan | Halodoc." Accessed: Jul. 04, 2024. [Online]. Available: https://www.halodoc.com/kesehatan/serangan-jantung

[7] A. Yang, W. Zhang, J. Wang, K. Yang, Y. Han, and L. Zhang, "Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA," *Front Bioeng Biotechnol*, vol. 8, 2020, doi: 10.3389/fbioe.2020.01032.

[8] J. Nasir, "Penerapan Data Mining Clustering dalam Mengelompokan Buku dengan Betode K-MEANS," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 2, pp. 690–703, Oct. 2021, doi: 10.24176/simet.v11i2.5482.

[9] J. J. Pangaribuan, J. Thames, A. E. Widjaja, O. P. Barus, and C. A. Haryani, "Classification of Dry Nut Types Using the Support Vector Machine Method," *2024 3rd International Conference on Creative Communication and Innovative Technology, ICCIT 2024*, 2024, doi: 10.1109/ICCIT62134.2024.10701219.

[10] D. A. Ryfai, N. Hidayat, and E. Santoso, "Klasifikasi Tingkat Resiko Serangan Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 10, pp. 4701–4707, Sep. 2022, Accessed: Jul. 04, 2024. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11662

[11] M. G. Pradana, P. H. Saputro, and D. P. Wijaya, "Komparasi Metode Support Vector Machine dan Naive Bayes dalam Klasifikasi Peluang Penyakit Serangan Jantung," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 2, p. 87, Dec. 2022, doi: 10.21927/ijubi.v5i2.2659.

[12] I. Optimasi *et al.*, "Implementasi Optimasi Hyperparameter GridSearchCV pada Sistem Prediksi Serangan Jantung Menggunakan SVM," *Teknologi: Jurnal Ilmiah Sistem Informasi*, vol. 13, no. 1, pp. 8–15, Jan. 2023, doi: 10.26594/TEKNOLOGI.V13I1.3098.

[13] C. Susanto, T. Taufiq, E. Hasmin, and K. Aryasa, "Sistem Pakar Prediksi Penyakit Diabetes Menggunakan Metode K-NN Berbasis Android," *CogITo Smart Journal*, vol. 8, no. 2, pp. 359–370, Dec. 2022, doi: 10.31154/COGITO.V8I2.406.359-370.

[14] M. N. A. H. Sha'abani, N. Fuad, N. Jamal, and M. F. Ismail, "kNN and SVM Classification for EEG: A Review," in *5th International Conference on Electrical, Control & Computer Engineering*, vol. 632, Singapore: Springer, 2020, pp. 555–565. doi: 10.1007/978-981-15-2317-5_47.

[15] B. Valentino Jayadi, T. Handhayani, and M. Dolok Lauro, "Perbandingan kNN dan SVM untuk Klasifikasi Kualitas Udara di Jakarta," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 11, no. 2, Aug. 2023, doi: 10.24912/jiksi.v11i2.26006.

[16] S. Wiyono, D. S. Wibowo, M. F. Hidayatullah, and D. Dairoh, "Comparative Study of KNN, SVM and Decision Tree Algorithm for Student's Performance Prediction," *International Journal of Computing Science and Applied Mathematics*, vol. 6, no. 2, p. 50, Aug. 2020, doi: 10.12962/j24775401.v6i2.4360.

[17] K. Vijila Rani and S. Joseph Jawhar, "Lung Lesion Classification Scheme Using Optimization Techniques and Hybrid (KNN-SVM) Classifier," *IETE J Res*, vol. 68, no. 2, pp. 1485–1499, Mar. 2022, doi: 10.1080/03772063.2019.1654935.

[18] M. J. Al Dujaili, A. Ebrahimi-Moghadam, and A. Fatlawi, "Speech emotion recognition based on SVM and KNN classifications fusion," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1259–1264, Apr. 2021, doi: 10.11591/ijece.v11i2.pp1259-1264.

[19] A. S. Remya Ajai and S. Gopalan, "Analysis of Active Contours Without Edge-Based Segmentation Technique for Brain Tumor Classification Using SVM and KNN Classifiers," in *Advances in Communication Systems and Networks*, vol. 656, Singapore:

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5719

Springer, 2020, pp. 1–10. doi: 10.1007/978-981-15-3992-3_1.

[20] R. Rahman, "Heart Attack Analysis & Prediction Dataset," 2021. Accessed: Jul. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

[21] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.

[22] A. Nazábal, C. K. I. Williams, G. Colavizza, C. R. Smith, and A. Williams, "Data Engineering for Data Analytics: A Classification of the Issues, and Case Studies," *ArXiv*, vol. abs/2004.12929, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:216552958