Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 33-40
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Analysis of Imbalanced Data in Heart Disease Classification Using SVM and SMOTE

Niken Ayunda*, Fadli Alfariddz Sinuraya, Fatika Fajar Akhfa Sinaga, Muhammad Rizieq Destian, Agus Perdana Windarto

Information Systems Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Email: [1,*]kenyunda@gmail.com, [2]fadlisinuraya12@gmail.com, [3]fatikasinaga@gmail.com, [4]riziqdestian@gmail.com, [5]agus.perdana@amiktunasbangsa.ac.id
Correspondence Author Email: kenyunda@email.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

Abstract−This study aims to examine the effect of applying the Synthetic Minority Over-sampling Technique (SMOTE) on the performance of the Support Vector Machine (SVM) algorithm in classifying heart disease in a dataset with class imbalance. The dataset used is derived from the 2015 BRFSS Heart Disease Health Indicators, where the number of non-patients far outweighs that of heart disease patients. This condition has the potential to reduce the model's ability to detect the minority class. Therefore, SMOTE was applied as a preprocessing step to balance the training data distribution before the classification process using SVM. Model evaluation was conducted using the accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results of the study indicate that the application of SMOTE is capable of improving the performance of the SVM model in detecting heart disease patients. The accuracy score increased from 0.77 to 0.78, recall from 0.78 to 0.83, F1-score from 0.75 to 0.76, and ROC-AUC from 0.84 to 0.85. Although there was a slight decrease in precision, the model's improved ability to recognize the minority class indicates that SMOTE is effective in addressing data imbalance issues. These results demonstrate that the combination of SVM and SMOTE can serve as a reliable approach for classifying heart disease in imbalanced datasets.

Keywords: Support Vector Machine; SMOTE; Heart Disease; Imbalanced Data; Classification

1. INTRODUCTION

The heart is a vital organ that pumps blood to supply oxygen and nutrients to the entire body (Galih et al., 2022). Heart disease is a condition in which vital components of the heart (such as the muscle, valves, or blood vessels) fail to function properly. This condition can be triggered by many factors, including arterial blockages, infections, inflammation, or congenital abnormalities(Sunyoto & Fatta, 2023). Heart disease not only affects individuals physically but also has social and economic implications, including an increased burden on the global healthcare system. According to WHO data, heart disease is one of the leading causes of global mortality, accounting for 7.4 million deaths in 2017 (Muzakki et al., 2024). In Indonesia, data from the Ministry of Health indicate that heart disease is the leading cause of death, with prevalence continuing to rise alongside changes in modern lifestyles, urbanization, and an aging population. Major risk factors include a diet high in saturated fat, lack of physical activity, stress, and air pollution, which exacerbate conditions such as coronary heart disease (CHD) and congestive heart failure (Nasution et al., 2025)-(Triyono et al., 2022). The impact is not limited to mortality but also includes morbidity, where patients often experience a decline in quality of life, disability, and high medical costs, thus requiring early intervention to prevent complications such as stroke or fatal heart attacks (Arifudin et al., 2023)-(Arista Kurniasari Budi Fristiani, 2025). Early classification of these diseases through the analysis of medical data, such as physical examination results, medical history, and biomarker parameters, is crucial for supporting accurate diagnosis, timely treatment, and the prevention of serious complications (Ngalle & Takalar, 2024)-(Supiyan, 2025). In the era of healthcare *big data*, medical datasets often originate from *electronic health records (EHRs),* which include thousands of attributes. Medical datasets frequently face class imbalance issues, where the number of heart disease patients is far fewer than healthy patients, hindering the development of effective classification models (Fitri et al., 2024)-(Komet Rama Daud, Parluhutan Sagala, Sutarno, n.d.)-(Melyani,Lensi Natalia Tambunan, 2023)-(Mariana Dewi, Triando Hamanongan Saragih, Rudy Herteno, Radityo Adi Nugroho, 2021). To address this, machine learning employs methods such as Support Vector Machine (SVM) and the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples for the minority class to make the data more balanced (Hermaliani & Ernawati, 2024). SVM is a supervised learning algorithm that works by finding the optimal hyperplane to separate data into different classes, focusing on the maximum margin between classes. SVM has the advantage of high generalization ability, but its weakness is sensitivity to imbalanced data, making it prone to bias toward the majority class and resulting in low accuracy for heart disease patients (Meisya et al., 2021)-(Zainul Arifin, Dhimas Fachri Rahman, Bagus Setya Rintyarna, 2023)-(Prabowo & Kurniadi, 2023). On the other hand, SMOTE is an oversampling technique that generates synthetic samples for the minority class by interpolating between existing samples, thereby improving the representation of the minority class without simply duplicating the data (Sofian Sidiq, Alfian, 2025)-(Muhammad Ibnu Choldun Rachmatullah, Sari Armiati, 2025). This technique is effective in balancing data distribution, but carries the risk of overfitting if the synthetic samples are too similar to the original data or do not reflect real-world variation, which can reduce model generalization (Salam et al., 2025).

Previous research has demonstrated the effectiveness of combining SVM and SMOTE across various domains, including healthcare and sentiment analysis. For example, (Nurrahman et al., 2025) shows that the SVM method is effective in analyzing public sentiment toward fuel price hikes; however, imbalanced data made it difficult for the model

to identify positive sentiment. With the application of SMOTE, the data distribution becomes more balanced, thereby increasing the model's accuracy to 84%. The research results also indicate that public sentiment tends to be negative, so the combination of SVM and SMOTE has proven suitable for sentiment analysis on datasets dominated by a single class (Nurrahman et al., 2025). Furthermore, research by Farini Nurhaliza et al., (2025) demonstrates that the use of SMOTE effectively addresses data imbalance in obesity classification, and when combined with a Linear kernel SVM, the model achieves the highest accuracy of 96.54%, surpassing both RBF and RUS. These findings confirm that data balancing techniques are crucial for improving the accuracy and fairness of predictions in health datasets (Martha, 2025). Additionally, research from (Indra et al., 2023) This study implemented sentiment analysis on Shopee customer reviews using the Support Vector Machine (SVM) method combined with the SMOTE-Tomek Links data balancing technique. This combination of methods proved superior, significantly improving classification performance, yielding an accuracy of 0.92 and an ROC AUC score of 0.97. This performance improvement indicates that data balancing successfully addressed the class imbalance issue in the review dataset. (Indra et al., 2023). The application of Support Vector Machines (SVMs) without special handling of imbalanced data, such as in heart disease diagnosis cases, often results in poor classification performance for the minority class and leads to high false negatives, thereby reducing medical reliability (Islam & Agung, 2025)-(Nurul et al., 2025)-(Herianto Siregar, Arifah Devi Fitriani, Aida Fitria, Ismail Efendy, 2024). Therefore, the proposed solution to improve recall and F1-score for the minority class is to use SMOTE as a preprocessing step to balance the data through synthetic sample generation, followed by training an SVM on the balanced data, where parameters such as kernel selection RBF and hyperparameters are optimized using techniques like grid search (Limbong, 2024)-(Agung Widyanto, Kusrini, 2023)-(Kumalasari et al., 2024). This approach not only addresses data imbalance but also ensures the model is more robust and reliable in medical applications..

This study aims to conduct an in-depth analysis of the synergistic effectiveness of the SVM algorithm enhanced with the SMOTE technique in improving the quality of heart disease classification facing the challenge of data imbalance. Results focused on improvements in accuracy, sensitivity, and specificity metrics will serve as the foundation for developing practical guidelines for AI applications in healthcare, significantly supporting the creation of more reliable clinical diagnostic tools.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study focuses on the analysis of imbalanced data in heart disease classification. In the methodology used, class imbalance is addressed through the SMOTE (Synthetic Minority Over-sampling Technique) oversampling technique. Furthermore, SVM (Support Vector Machine) is implemented as the classification algorithm. The entire analysis process is visualized through a flowchart.

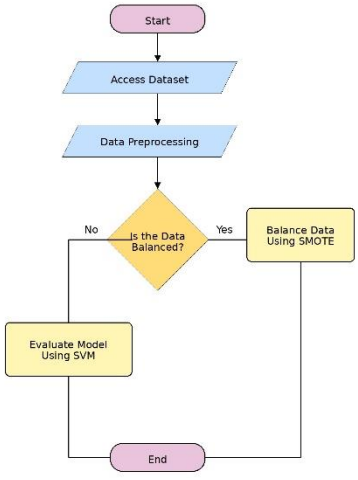

Figure 1. Research workflow

Figure 1 summarizes the main steps in handling imbalanced data for heart disease classification using SMOTE and SVM. The process begins with data collection and preprocessing. The next crucial step is checking class balance; if the data is found to be unbalanced, SMOTE is implemented to oversample the minority class. The balanced dataset then undergoes training and testing using the SVM algorithm. The entire analysis concludes after the classification model's performance evaluation is completed.

2.2 Dataset

Table 1 Shows that this study uses a secondary dataset obtained from Kaggle, namely the Heart Disease Health Indicators BRFSS 2015. This dataset contains data from a public health survey that includes various indicators related to heart

disease risk. The target variable in this study is heart disease status, specifically whether an individual has been diagnosed with heart disease or not. Meanwhile, other variables consist of health indicators such as blood pressure, cholesterol, body mass index (BMI), smoking habits, physical activity, health status, and age, which are used to help predict the risk of heart disease.

Table 1. Dataset

High BP	High Col	Age	Bmi	Smoker	Phys Activity	MentHlth	HvyAlcohol Consump	Veggies	PhysHlth
1	1	1	40	1	0	18	0	1	15
0	0	0	25	1	1	0	0	0	0
1	1	1	28	0	0	30	0	0	30
1	0	0	27	0	1	0	0	1	0
1	1	1	24	0	1	3	0	1	0
….	….	….	….	….	….	….	….	….	….
1	0	8	33	0	1	0	0	1	0

2.3 *Prepocessing*

The preprocessing stage is crucial for preparing the dataset prior to the core analysis and classification process (Andriyani et al., 2025). In this step, a series of activities are performed, including the identification and handling of duplicate data and missing values, and the adjustment of the data format to ensure compatibility with the model. Additionally, feature normalization or scaling is applied to standardize the range of variable values, which is essential for optimizing the performance of the SVM algorithm. This stage also involves splitting the dataset into training and testing data to ensure the objectivity of the model evaluation. Preprocessing in this study includes data preprocessing and feature preprocessing:

2.3.1 *Data Preprocessing*

In Table 2, the data used has undergone a cleaning process to ensure there are no missing values and all attributes have a uniform format. The data was then prepared for effective use in the modeling process. In this stage, the target label HeartDisease or Attack was also defined, indicating the respondent's heart disease status, with a value of 1 for those with heart disease and 0 for those without. The result of this stage is cleaner data ready for analysis and classification.

Table 2. Data after preprocessing

HighBP	HighChol	BMI	Smoker	Phys Activity	Age	PhysHlth	MentHlth	HvyAlcohol Consump	Sex
1	1	40	1	0	9	15	18	0	0
0	0	25	1	1	7	0	0	0	0
1	1	28	0	0	9	30	30	0	0
1	0	27	0	1	11	0	0	0	0
1	1	24	0	1	11	0	3	0	0
1	1	25	1	1	10	2	0	0	1
1	0	30	1	0	9	14	0	0	0
….	….	….	….	….	….	….	….	….	….
1	1	33	1	1	12	2	0	0	1

This stage involves a series of fundamental procedures carried out to ensure the quality and readiness of the dataset before it enters the analysis process. These procedures include:

a. Duplicate Removal: Duplicate data is identified and removed to ensure that observations in the dataset remain independent, avoiding unnecessary bias.
b. Handling Missing Values: Missing values in the dataset are detected and addressed, either through imputation methods (such as using the mean or median) or by removing rows of data, depending on the findings.
c. Data Type Standardization: All features are converted to numeric data types. This is an important prerequisite for the SVM algorithm to process the data effectively.

2.3.2 *Feature Preprocessing*

The feature preprocessing stage is carried out to condition the predictor variables. The goal is to ensure that each feature has a uniform and optimal range of values and distribution before being fed into the classification process.

a. Data Normalization/Standardization
   This stage is an essential step because the SVM algorithm is highly sensitive to differences in feature scale. Therefore, normalization is performed to ensure that the value ranges of each feature are uniform, guaranteeing an equal contribution to the model's distance calculation. Two commonly used techniques are:
b. Min–Max Normalization

Min-Max normalization is a data scaling technique that transforms a feature's values into a defined range, typically [0, 1]. This process is achieved by mapping each original data value(X) based on the minimum($Xmin$) and maximum($Xmax$) values present in that feature.min

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

c. Standardization (Z-Score Normalization)
Z-Score Normalization (Standardization) transforms feature values such that the data has a mean of 0 and a standard deviation of 1. This technique is an effective choice, particularly when dealing with data that has an uneven distribution.

$$X^I = \frac{X - \mu}{\sigma} \quad (2)$$

In this process, each original data value (X) is subtracted by the feature's mean (μ), and the result is then divided by the standard deviation (σ). This transformation yields a new value (X') with a mean close to 0 and a standard deviation of 1. Consequently, scale differences between features are minimized, making the data more balanced and easier for machine learning algorithms to process.

d. Dataset Split (Train-Test Split)
The dataset is divided into two parts with a specific ratio (e.g., 80:20), where one part is used to train the model and the other is used to test the model's predictive capabilities. This split is a mandatory step to ensure the objectivity of the evaluation of the trained model's perfomance.

$$Train = p \; x \; N \quad (3)$$

$$Test = (1 - p) \; x \; N \quad (4)$$

80% of the total data (N) is used as training data to train the model, while the remaining 20% is used as test data to evaluate the model's performance. The amount of training data is calculated using the equation Train = p × N, while the amount of test data is calculated using the equation Test = (1 − p) × N, where p represents the training data proportion of 0.8.

2.4 *Support Vector Machine*

The Support Vector Machine (SVM) algorithm is a supervised learning technique used to classify data and recognize patterns within it. SVM works by linearly separating classes using a hyperplane. To address non-linear problems, SVM utilizes the concept of kernel functions (the kernel trick). The basic SVM equation can be expressed as follows**:**

$$f(x) = w^T \text{x} + \text{b} \quad (5)$$

In the basic Support Vector Machine (SVM) equation, `wT ` represents the transpose of the weight vector that defines the orientation of the separating hyperplane. Meanwhile, `x ` is the input vector representing each data point to be classified, and `$b$ ` is the bias or constant used to shift the position of the hyperplane.

$$[(w^T . x_i + \text{b})] \geq 1 \text{ untuk } y_i = +1 \quad (6)$$

$$[(w^T . x_i + \text{b})] \leq -1 \text{ untuk } y_i = -1 \quad (7)$$

Where `xi ` is the i-th input data vector, and `yi ` is the corresponding class label. The optimal hyperplane in SVM is obtained by optimizing the distance between two groups of objects from different classes. This maximum distance is called the margin. The margin is calculated using a formula involving the norm (length or magnitude) of the weight vector `w`.

2.5. SMOTE

SMOTE (Synthetic Minority Oversampling Technique) is a commonly used method for addressing class imbalance issues in data. According to the Google Machine Learning Crash Course, data imbalance can be classified based on the proportion of the minority class: mild imbalance when the minority class is in the range of 20–40%, moderate imbalance in the range of 1–20%, and extreme imbalance if the proportion is less than 1%

$$x_{syn} = x_i + (x_{knn} - x_i)\text{y} \quad (8)$$

In the SMOTE equation, `$xsyn$ ` represents the new synthetic data generated to increase the sample size of the minority class. This synthetic data is formed based on the original minority class samples denoted as `$xi$ `. During this process, the algorithm searches for another sample with the most similar characteristics or the nearest neighbor ` ` of `$xi$ `, denoted as `$xknn$ `. Next, new data is generated at a specific position betweenxi and$xknn$, so that the characteristics of the synthetic data remain within the distribution space of the minority class. The variable (Y) indicates the label or class of each data point, for example class 0 and class 1, which are used to distinguish data categories in the classification process.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 33-40
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

3. RESULTS AND DISCUSSION

This section presents the results of the Support Vector Machine (SVM) model evaluation in heart disease classification through scenarios before and after the implementation of the SMOTE technique. The analysis was conducted descriptively and quantitatively to measure the significance of data balancing on model performance optimization, particularly regarding detection accuracy for heart disease patient categories.

3.1 Data Description and Initial Conditions (Imbalanced Data)

In the initial stage of the research, an analysis was conducted on the dataset used to determine the characteristics and distribution of data in each class.

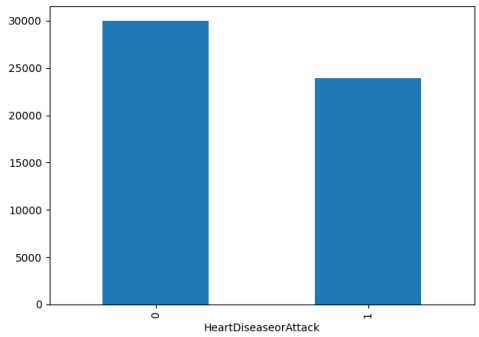

Figure 2. Imbalanced Data Distribution in the Heart Disease Dataset

Figure 2 illustrates the significant disparity in sample sizes between Class 0 and Class 1 in the heart disease dataset. The dominance of the majority class poses a challenge to classification performance, as the model may experience a decrease in sensitivity when detecting heart disease patients. Therefore, the application of data balancing methods is necessary to ensure the model can learn the characteristics of both classes objectively

3.2 Training Data Distribution Before and After SMOTE

An analysis of the training data distribution was conducted before and after applying SMOTE to observe changes in the proportions between classes. This analysis aims to evaluate the level of data balance achieved after the oversampling process.

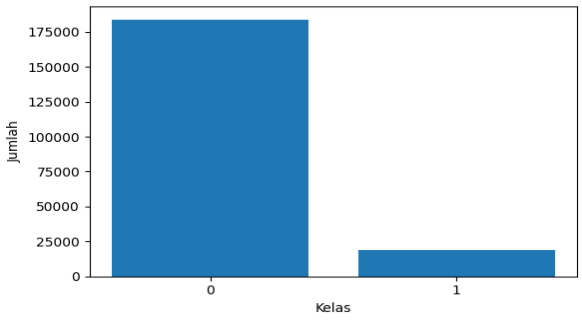

Figure 3. Training Data Distribution Before SMOTE Application

Figure 3 shows that the amount of data in the non-heart disease class is much greater than in the patient class. This imbalance can cause the model to be more inclined to predict the majority class, so further handling is required before the modeling process is carried out.

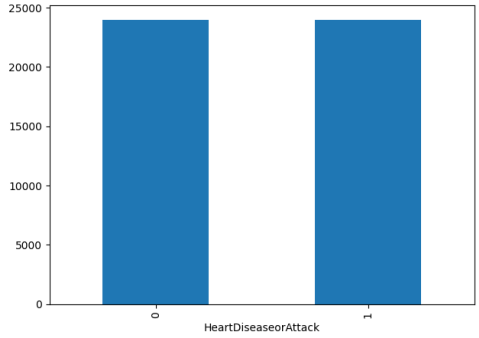

Figure 4. Training Data Distribution After Applying SMOTE

Figure 4 shows that the application of SMOTE results in a more balanced distribution of training data between the heart disease and non-heart disease classes, making the data representation for model training more proportional.

3.3 Model Evaluation Using F1-Score and ROC-AUC

Model performance evaluation was conducted using the F1-Score and ROC-AUC metrics. The F1-Score is used to measure the balance between precision and recall, while the ROC-AUC is used to assess the model's ability to distinguish between positive and negative classes. Higher values for both metrics indicate better model performance in classification.

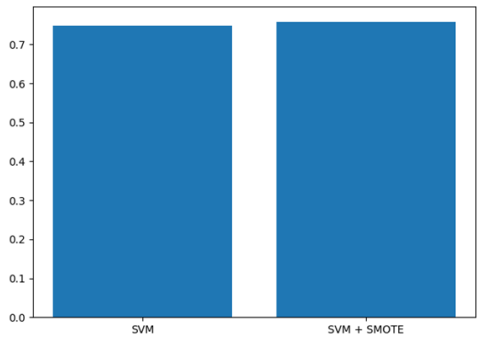

Figure 5. Comparison of F1-Score Values for the SVM and SVM+SMOTE Models

Figure 5. The comparison in the figure shows that the use of SMOTE successfully boosted the F1-score of the SVM algorithm. This improvement reflects the model's success in achieving a balance between precision and recall, resulting in a substantial improvement in detection performance for the category of heart disease patients compared to before

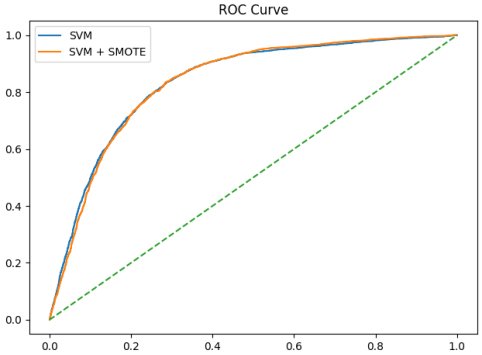

Figure 6. ROC Curves for the SVM and SVM+SMOTE Models

Figure 6. presents the ROC curve as an indicator of the model's effectiveness in distinguishing between the two target classes. The significant AUC value indicates that the model is not only capable of accurately separating patient classes but also demonstrates strong performance stability across various classification thresholds.

3.4 Comparison of Model Performance Before and After SMOTE

In Table 3, applying SMOTE to the SVM algorithm successfully improved the accuracy, recall, F1-score, and ROC-AUC values compared to SVM without SMOTE. The increase in recall from 0.78 to 0.83 indicates that the model is more effective at identifying the minority class, while the rise in ROC-AUC from 0.84 to 0.85 suggests improved class separation capability. Although precision saw a slight decrease, overall, the use of SMOTE successfully enhanced the SVM model's classification performance on imbalanced data.

Table 3. Comparison of SVM Model Performance

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
SVM	0.77	0.72	0.78	0.75	0.84
SVM+ SMOTE	0.78	0.70	0.83	0.76	0.85

3.5 Discussion

This study demonstrates that the use of SMOTE in the SVM model yields positive results. This study analyzes the impact of integrating the SMOTE technique on the performance of the Support Vector Machine (SVM) in classifying heart disease on an imbalanced dataset. Initially, the significant imbalance in the proportion of patients and non-patients risked causing bias toward the majority class. The implementation of SMOTE on the training data successfully balanced the distribution of both classes, allowing the model to extract patterns in the minority class more effectively. Test results showed a substantial increase in recall, from 0.78 to 0.8306, representing an enhancement in the model's sensitivity for

identifying patients with heart disease. In the medical field, this improvement is crucial for minimizing the risk of false negatives. Although there is a trade-off in the form of a decrease in precision to 0.6984, the application of SMOTE also improves overall evaluation metrics, as demonstrated by an increase in accuracy to 0.78 and an ROC-AUC value reaching 0.85. It can be concluded that SMOTE effectively enhances the ability to detect positive cases without degrading overall class separation performance.

4. CONCLUSION

Based on the research results, it can be concluded that the application of SMOTE to the Support Vector Machine (SVM) algorithm is effective in addressing the issue of imbalanced data in heart disease classification. SMOTE is capable of balancing class distributions so that the model is more sensitive in detecting patients with heart disease. The increase in recall and F1-score values indicates that the model has become more reliable in identifying positive cases, despite a slight decrease in precision. The relatively stable accuracy and ROC-AUC values indicate that the application of SMOTE does not reduce the model's overall discriminatory ability. Thus, the combination of SVM and SMOTE is highly recommended for medical classification cases with imbalanced data, especially when the primary goal is to minimize errors in detecting patients who actually have the disease.

REFERENCES

Agung Widyanto, Kusrini, K. (2023). Pengaruh Keseimbangan Data Terhadap Akurasi Model Support Vector Machine Pada Data Set Donor Darah. *Jurnal Teknologi Terpadu*, *9*(2), 79–88.

Andriyani, D., Faqih, A., & Permana, S. E. (2025). *The Effect of SMOTE Application on Support Vector Machine Performance in Sentiment Classification on Imbalanced Datasets*. *4*(2).

Arifudin, N. F., Kristinawati, B., Studi, P., Fakultas, K., Kesehatan, I., Muhammadiyah, U., Jakarta, S., Studi, P., Fakultas, K., Kesehatan, I., Muhammadiyah, U., Jakarta, S., & Reviews, S. (2023). *HIJP : HEALTH INFORMATION JURNAL PENELITIAN Dampak Masalah Psikologis Terhadap Kualitas Hidup Pasien Gagal Jantung: Systematic Review*. *15*.

Arista Kurniasari Budi Fristiani, G. M. P. (2025). *Skrining Hipertensi Sebagai Langkah Awal Pencegahan Penyakit Jantung dan Stroke*. *4*(1), 11–15.

Fitri, A., Masruriyah, N., Novita, H. Y., Sukmawati, C. E., & Ramadhan, A. R. (2024). *Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung*. *4*(1), 62–70.

Galih, D., Luthfi, M., & Farhan, M. (2022). *Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network*. *3*(2), 55–60.

Herianto Siregar, Arifah Devi Fitriani, Aida Fitria, Ismail Efendy, N. (2024). Analisis Implementasi Sistem Informasi Rumah Sakit Terhadap Pelayanan Administrasi Rumah Sakit Haji Syaiful Anwar. *Jurnal Promotif Preventif*, *7*(5), 1011–1021.

Hermaliani, E. H., & Ernawati, M. (2024). Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian. *Computer Science (CO-SCIENCE)*, *4*(1), 80–88.

Indra, D., Endro, J., & Amien, W. (2023). *Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction*. *01*, 1–9. https://doi.org/10.21456/vol13iss1pp1-9

Islam, U., & Agung, S. (2025). Implementasi Teknik Resampling Untuk Mengatasi Ketidakseimbangan Data Terhadap Klasifikasi Anemia Menggunakan Support Vector Machine. *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, *2*(3), 942–951.

Komet Rama Daud, Parluhutan Sagala, Sutarno, S. (n.d.). *Analisis Yuridis Hukum Rekam Medis Elektronik Sebagai Alat Bukti Dalam Suatu Sengketa Medis*.

Kumalasari, N. O. R., Pratiwi, C., & Ibrahim, N. U. R. (2024). *Prediksi Kanker Paru menggunakan Grid search untuk Optimasi Hyperparameter pada Algoritma MLP dan Logistic Regression*. *12*(3), 556–568.

Limbong, H. H. (2024). *Optimasi Analisis Sentimen Ulasan Aplikasi Amikom One Menggunakan SMOTE pada Algoritma Artificial Neural Network Optimization of Sentiment Analysis for Amikom One Application Reviews Using SMOTE with Artificial Neural Network Algorithm*. *13*, 2048–2059.

Mariana Dewi, Triando Hamanongan Saragih, Rudy Herteno, Radityo Adi Nugroho, D. T. N. (2021). Penerapan SMOTE-NCL Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Penyakit Jantung Koroner. *JIP (Jurnal Informatika Polinema)*, *10*(1), 27–33.

Martha, S. (2025). Penerapan Suport Vector Machine dengan Synthetic Minority Oversampling Technique (Smote) Dalam Pengklasifikasi Penyakit Diabetes. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, *14*(3), 262–271.

Meisya, T., Aulia, P., Arifin, N., Mayasari, R., Informatika, T., Komputer, F. I., & Karawang, U. S. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *SINTECH JOURNAL*, *4*(2), 139–145.

Melyani,Lensi Natalia Tambunan, E. P. B. (2023). Hubungan Usia dengan Kejadian Penyakit Jantung Koroner pada Pasien Rawat Jalan di RSUD dr. Doris Sylvanus Provinsi Kalimantan Tengah. *Jurnal Surya Medika (JSM)*, *9*(1), 119–125.

Muhammad Ibnu Choldun Rachmatullah, Sari Armiati, M. (2025). *Menerapkan smote pada klasifikasi data penyakit stroke*. *17*(1), 9–12.

Muzakki, F., Ubaydillah, I., Assyiami, N. R., & Soleha, S. (2024). *Penerapan Algoritma C4 . 5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer*. *2*, 71–79.

Nasution, I. S., Rahmadani, A. D., Audina, W., & Purnama, D. (2025). *Systematic Review : Pengaruh Gaya Hidup dan Pengetahuan Masyarakat terhadap Risiko Penyakit Jantung Koroner*. *4*(2), 287–298. https://doi.org/10.54259/sehatrakyat.v4i2.4337

Ngalle, P., & Takalar, K. (2024). *Arus Jurnal Sains dan Teknologi (AJST) Analisa Diagnosa Penyakit Berdasarkan Riwayat Medis menggunakan Algoritma Random Forest Studi Kasus Rumah Sakit*. *2*(2).

Nurrahman, A. A., Mauladi, M., & Rahman, A. (2025). *Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE*. *90*.

Nurul, L., Pulungan, T., & Utami, E. (2025). *Perbandingan algoritma svm dan random forest pada klasifikasi kesehatan mental berdasarkan sleep disorders*. *10*(3), 2827–2837.

Prabowo, A. S., & Kurniadi, F. I. (2023). *Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung*.

Salam, A., Azhari, L., Septarini, R. S., & Heriyani, N. (2025). *BULLETIN OF COMPUTER SCIENCE RESEARCH Pendekatan Hybrid K-Means SMOTE dan Logistic Regression Untuk Deteksi Dini Diabetes Mellitus Pada Imbalanced Data*. *5*(3), 219–227. https://doi.org/10.47065/bulletincsr.v5i3.502

Sofian Sidiq, Alfian, N. S. M. (2025). Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling. *JURNAL ILMIAH INFORMATIKA DAN ILMU KOMPUTER (JIMA-ILKOM)*, *4*(1), 13–23.

Sunyoto, A., & Fatta, H. Al. (2023). *Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier*. *VII*(September), 31–40.

Supiyan, D. (2025). *Pengembangan Sistem Pakar Untuk Diagnosa Penyakit Diabetes Melitus Menggunakan Metode Forward Chaining*. *7*(3). https://doi.org/10.32877/bt.v7i3.2244

Triyono, D., Liani, R., Utami, A. W., Tristiyanti, S., Supriatna, A., Surabaya, K. P., & Bandung, S. B. (2022). *PENYAKIT JANTUNG KORONER DI INDONESIA : PERAN*. *17*(1), 86–94.

Xgboost, D. A. N., & Prediksi, U. (2025). Perbandingan Algoritma Machine Learning Svm,Random Forest,dan Xgboost Untuk Prediksi Stroke. *Jurnal Teknologi Dan Sistem Informasi Univrab*, *10*(2), 1098–1110.

Zainul Arifin, Dhimas Fachri Rahman, Bagus Setya Rintyarna, D. (2023). Penerapan Algoritma Support Vector Machine Berbasis Kernel Radial Basis Function dalam Klasifikasi Sel Kanker. *BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer*, *4*(2), 100–106.