**Jurnal Informasi dan Teknologi**
https://jidt.org/jidt
2025 Vol. 7 No. 2 Hal: 94-101 e-ISSN: 2714-9730

# Precision Medicine Through Support Vector Machines: Analyzing Patient Data for Improved Drug Classification

Nanda Rosma Anwar[1]✉, Dendy K Pramudito[2], Muhammad Makmun Effendi[3]

[123]Universitas Pelita Bangsa

jurnal.jidt@gmail.com

**Abstract**

Selecting the appropriate medication is crucial for ensuring optimal therapeutic outcomes and minimizing adverse effects for patients. Healthcare personnel are managing an increasing volume of medical data in the digital era. Identifying swift, precise, and dependable methods for recommending appropriate medications is becoming essential. This study aims to meet this criterion by classifying drugs into appropriate categories for patient care using the Support Vector Machine (SVM) technology. The research utilized a dataset from GitHub comprising 200 patient records. These records furnish critical information regarding the patient, including age, sex, blood pressure, cholesterol levels, sodium-to-potassium ratios, and prescriptions. To maximize the use of this data, the method entails several critical steps: selecting appropriate data, meticulously cleaning and organizing it, transforming it for analytical readiness, employing SVM for data mining, and conducting a comprehensive review. The dataset is divided into two segments which are 20% is allocated for testing the efficacy of the SVM model, while the remaining 80% is designated for training the model.The primary tool for constructing the SVM model is the Google Colaboratory platform, which utilizes Python. A confusion matrix is employed to meticulously evaluate the performance of a model. It provides valuable metrics such as accuracy, precision, recall, and the F1 score. The evaluation method indicates that the SVM model holds significant potential for systematically assessing patient data due to its capability to appropriately categorize various drug types. This discovery represents a significant advancement for AI in healthcare, as it facilitates the prompt and straightforward recommendation of individualized medicines by physicians.

**Keywords:** *Support Vector Machine, Drug Classification, Data Mining, Machine Learning, Artificial Intelligence.*

## 1. Introduction

Technology is moving forward at an unprecedented speed in today's world, affecting almost every field, including healthcare. In healthcare, it is very important to choose the right prescription for a patient's condition. This demand exists because everyone has different health traits that are affected by things like age, gender, blood pressure, cholesterol levels, and electrolyte balance. These individual differences have a big impact on how well a patient responds to certain medications. Therefore, in order to give a patient the right medical care, you need to know a lot about their whole health history and carefully look at all of their symptoms. Choosing the wrong medicines can have serious health impacts, ranging from mild side effects to severe problems that could make a patient's condition worse. It is becoming harder to make smart choices because there is so much medical information out there and because individuals can have a wide range of symptoms [1]. These problems show how important it is to have advanced tools that can quickly and accurately assess patient data to help healthcare professionals make decisions. Modern medicines are an important part of daily life since they are used for more than just treating illnesses; they also help people stay healthy and avoid getting sick. Access to medication has greatly improved, since drugs may now be found in a variety of places, such as small booths, specialty drug stores, and certified pharmacies. Despite this improved accessibility, a large part of the population still doesn't know about the different types and groups of pharmaceuticals. The lack of knowledge is a big problem because pharmaceuticals are divided into several groups, including over-the-counter drugs, prescription drugs, controlled substances, opioids, anesthetics, and traditional or herbal remedies. Taking medications the wrong way, especially ones that require a prescription from a doctor, can be quite dangerous for your health. Using strong or limited drugs without enough medical monitoring can be bad for your health and, in some cases, even deadly. As a result, it is important right once to set up reliable methods for drug classification so that both the general public and healthcare professionals can accurately identify and use drugs. Using advancements in information technology provide a feasible solution for addressing these challenges. Machine learning algorithms are quite good at handling complicated classification tasks. The Support Vector Machine (SVM) stands out among these

Receipt: 02-01-2025 | Revision: 10-03-2025 | Publish: 14-07-2025 | doi: 10.60083/jidt.vi0.627

techniques for its effectiveness and accuracy in classifying data, especially medical data. Using the SVM algorithm in pharmacological classification systems can significantly improve the accuracy of drug identification, leading to safer and more successful therapeutic outcomes [2].

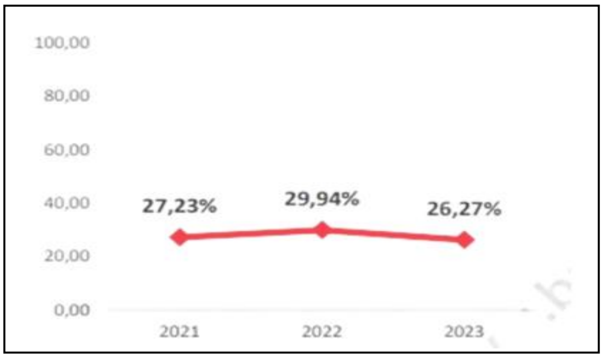


Figure 1. Percentage of Population with Health Complaints (2021-2023) [3]

There are advances in science and technology in the last three years, explaining the number of residents who have health complaints refer to figure 1. This data shoes that the highest percentage occurred in 2022 while 2023 showed a significant decrease. This could indicate the presence of certain factors, such as changes in environmental conditions, access to health services or other factors that affect public health [3]. Support Vector Machine (SVM) is one of yhe machine learning algorithms that is widely used in classification problems, mainly because of its ability to separate data from two or more classes using an optimal hyperlane. SVM can be used to classify drug types for patients. And able to handle high dimensional data [4]. In the context of drug type classification, SVM can process various patient medical parameters, such as blood pressure, cholesterol levels and sodium to potassium ratio, to help recommend the most appropriate drug type and show that SVM based models are able to achieve high performance in classifying drug for patients [5]. One of the main advantages of the Support Vector Machine (SVM) algorithm is the simplicity in implementing it and its abilityto handle complex datasets. SVM is very effective in various classification cases including in the health sector[6]. Previous research shows that K-Nearest Neighbor (KNN) algorithm can work well and determine the appropriate type of medication based on the patient's medical parameters, such as blood pressure, cholesterol levels and sodium to potassium ratio. The results of this study show that the value of k = 3 and the ratio of training and test data division of 90:10 resulted in an accuracy of 98,33%, precision 98,8%, recall of 96,2% and f-measure of 97,48%. And experienced a decrease in performance when using the cross validation model. By using the K-Nearest Neighbor (KNN) algorithm, it is hoped that it can help doctors to determine the right type of medicine for patients based on the patient's condition[1]. Based on the description above, the research conducted used the Support Vector Machine (SVM) algorithm. This research aims to classify drugs based on patient data. The main objective of this research is to improve efficiency and accuracy in providing appropriate treatment recommendarions. By utilizing the right dataset and algorithm, this Support Vector Machine based system for drug classification for patients is expected to help medical personnel make faster and more accurate decisions and the results of this study are expected to contribute to the development of more sophisticated health technology and support better media services in the future [7].

## 2. Research Methods

Quantitative method is used with secondary data sourced form the github platform for this research study.

### 2.1. Data Mining Workflow

The data mining workflow starts from problem identification, namely errors in drug selection based on patient symptoms. This study aims to classify drug types using the Support Vector Machine algorithm based on patient data. A datased of 200 data is taken from github. The next steps include attribute selection, pre-processing, transformation, data mining and model evaluation. In the data mining process, there is 80% training data and 20% testing data with testing with thr random sampling method based on the slovin formula. Model evaluation is done with confusion matrix to calculate accuracy, precision, recall and f1-score. Implemented with google collaboratory.

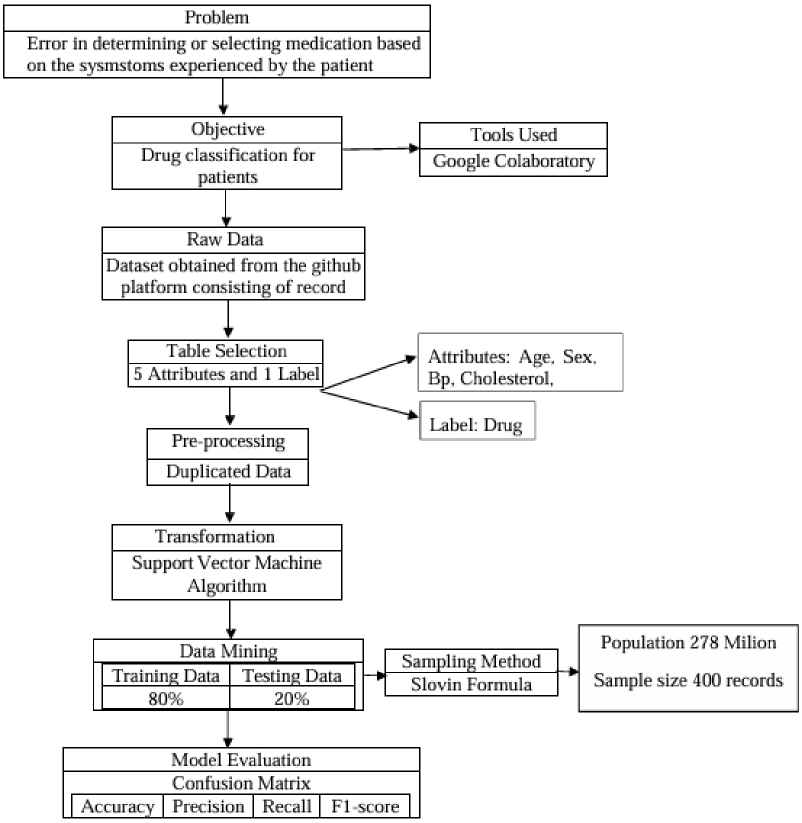


Figure 2. Data Mining Workflow

2.2. KDD (Knowledge Data in Discovery) Stages

This research method uses the KDD (Knowledge Data in Discovery) stage. Implementing the Support Vector Machine algorithm to classify drugs against patients.

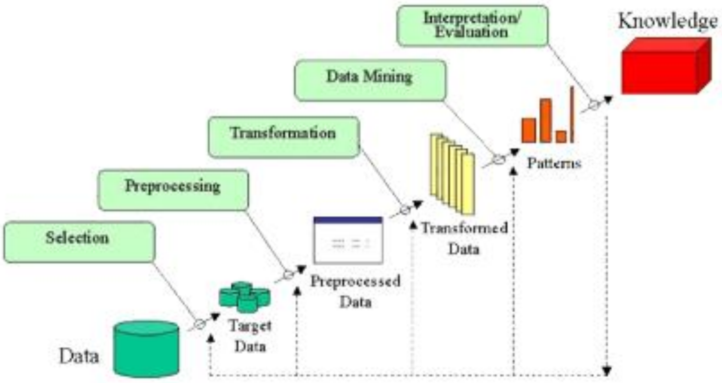


Figure 3. KDD Process Stage

a. Initial Data

The initial data used in this research is raw data obtained from various sources and has not undergone processing stages such as selection, pre-processing or futher analysis. This dataset amounts to 200 data, which consists of relevant attributes or variables in the form of Age, Sex, Bp, Cholesterol, Na_to_K and Drug.

Table 1. Drug Dataset

| No | Age | Sex | Bp | Cholesterol | Na_to_K | Drug |
|---|---|---|---|---|---|---|
| 1 | 23 | F | High | High | 25.355 | Drug Y |
| 2 | 47 | M | Low | High | 13.093 | Drug C |
| 3 | 47 | M | Low | High | 10.114 | Drug C |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | 28 | F | Normal | High | 7.798 | Drug X |
| 5 | 61 | F | Low | High | 18.043 | Drug Y |
| 6 | 22 | F | Normal | High | 8.607 | Drug X |
| 7 | 49 | F | Normal | High | 16.275 | Drug Y |
| 8 | 61 | M | Low | High | 11.037 | Drug C |
| 9 | 60 | M | Normal | High | 15.171 | Drug Y |
| 10 | 43 | M | Low | Normal | 19.368 | Drug Y |
| ….. | …… | ……. | ……… | ………… | ………….. | ……… |
| 200 | 40 | F | Low | Normal | 11.349 | Drug X |

b. Selection

In data selection, an understanding of the data to be processed is very important because the research data does not yet have their own attributes and variables before the data mining process begins.

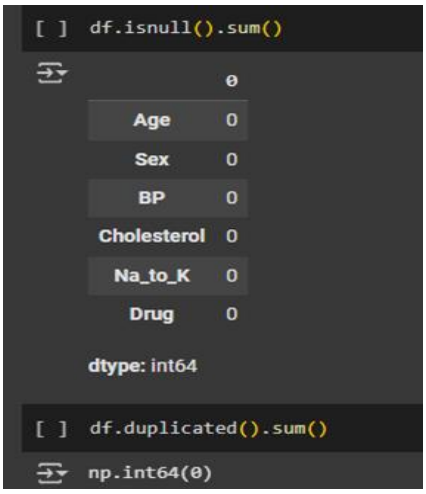
Table 2. Data Selection

| No | Attribute | Indicator | Usage Detail |
|---|---|---|---|
| 1 | Age | O | Model Value |
| 2 | Sex | O | Model Value |
| 3 | Bp | O | Model Value |
| 4 | Cholesterol | O | Model Value |
| 5 | Na_to_K | O | Model Value |
| 6 | Drug | O | Label |

The table above shows that the “o” indicator is the attribute to be used and the “x” indicator is the attribute to be eliminated in the initial data stage. In this stage, there are 5 attributes and 1 label.

c. Pre-processing

Data cleaning is done to fill or eliminate duplicate data and data that has no value so thet the data to be processed is clean and ready to use.



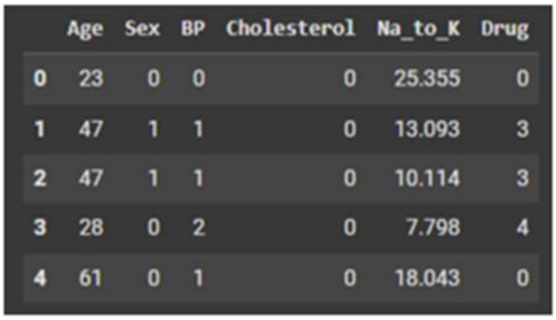
```
[ ] df.isnull().sum()

                 0
Age              0
Sex              0
BP               0
Cholesterol      0
Na_to_K          0
Drug             0

dtype: int64

[ ] df.duplicated().sum()

np.int64(0)
```

Figure 4. Duplicated Data

The picture shows that there is no duplicated data and missing values.

d. Transformation

At the next stage, it is done to change the source data to a different format so that it can be used by the data processing process at the next stage. In addition, attribute values must also be changed, which can cause the data pattern recognition an decision making process to take longer.



| | Age | Sex | BP | Cholesterol | Na_to_K | Drug |
|---|---|---|---|---|---|---|
| 0 | 23 | 0 | 0 | 0 | 25.355 | 0 |
| 1 | 47 | 1 | 1 | 0 | 13.093 | 3 |
| 2 | 47 | 1 | 1 | 0 | 10.114 | 3 |
| 3 | 28 | 0 | 2 | 0 | 7.798 | 4 |
| 4 | 61 | 0 | 1 | 0 | 18.043 | 0 |

Figure 5. Transformation Data

With the transformation using LabelEncoder, to convert categorical data info numerical data that can be used in machine learning algorithms.

e. Data Mining
At this stage, data mining used to classify drugs against patients. The dataset consist of 200 data, divided into two data, namely 80% training data and 20% testing data, according to the standard data division in the machine learning process. Calculation of training data and testing data as follow.

Total data (N) = 200

Total training data = 80% × 200 = 160

Total testing data = 20% × 200 = 40

After the data calculation is complete and in accordance with the previous processing stages. Then the processing is carried out using the python based google colaboratory platform.

Sample size

This research was conducted to obtain the condition of the population is very large and cannot be known with certainty, the slovin formula is used to determine the minimum sample size. This formula is used to determine the minimum sample size. This formula was chosen because it can help calculate the sample size by considering the predetermined margin of error.

$$n = \frac{N}{1+N.e^2}$$

Where:

n = Minimum required size

N = Population size

$e^2$ = Margin of error (in decimal form)

$$n = \frac{278.000.000}{1+278.000.000 \times 0,005^2}$$

$$n = 400$$

The amount of data aivable in this study is 200 data, which is still below the recommended minimum sample size. Based on the estimated population of 278.000.000, this number is not enough to produce an optimal analysis. Therefore, the slovin formula was used to calculate the required sample size. With a confidence level of 95% and a margin of error of 5%, the ideal minimum sample size for this study is 400 data. This sample size is considered sufficient to represent the population thoroughly and produce a more accurate sample.

f. Evaluation
At this stage, the evaluation process is carried out by observing and analyzing the performance of the algorithm to ensure the suitability of the test results with the research objectives. Meanwhile, validation aims to measure the model’s ability to make accurate predictions. Therefore, accuracy is used as the main evaluation matrix to assess the performance of the applied algorithm.

## 3. Results and Discussion

### 3.1. Confusion Matrix

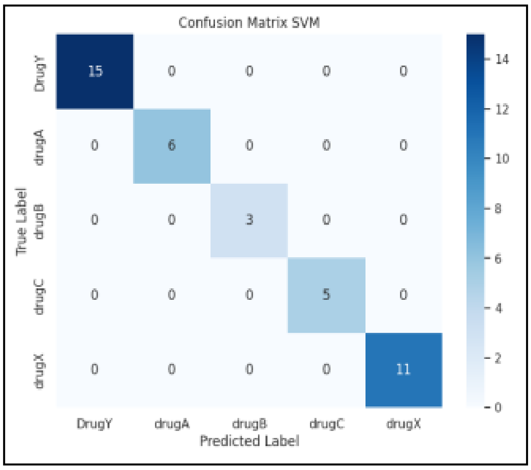
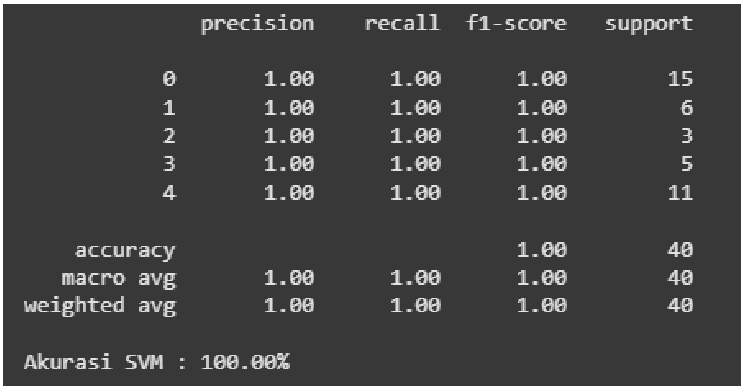


Figure 6. Confusion Matrix

This confusion matrix shows the prediction results of the SVM model against the testing data. DrugY was predicted correctly for 15 patients, who generally had high blood pressure (Bp) and low (Na_to_K) ratio. DrugA was suitable for 6 patients and all were correctly predicted. DrugB was administered to 3 patients and was correctly classified. DrugC was assigned to 5 patients and all of them were also accurately predicted. Meanwhile, DrugX was given to 11 patients who were in stable condition and was classified correctly. Overall, the SVM model successfully classified the entire dataset without error and it is evident from all the values on the main diagonal (True Positive) that the model has very good prediction accuracy.

### 3.2. Evaluation



```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        15
           1       1.00      1.00      1.00         6
           2       1.00      1.00      1.00         3
           3       1.00      1.00      1.00         5
           4       1.00      1.00      1.00        11

    accuracy                           1.00        40
   macro avg       1.00      1.00      1.00        40
weighted avg       1.00      1.00      1.00        40

Akurasi SVM : 100.00%
```

Figure 7. SVM Evaluation

The evaluation stage, using the Support Vector Machine (SVM) algorithm, obtained precision, recall and f1-score which is worth 100% in all classes. With an accuracy of 100%, which shows the model's performance is very good in classification. And this SVM model is able to predict all tesying data correctly and without error.

The calculation of the Support Vector Machine algorithm:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total test data}} = \frac{40}{40} = 100\%$$

The support is 40 (15 + 6 + 3 + 5 + 11) and since all predictions are correct, the accuracy is 100%. From this data, we get:

True Positive (TP) = 15

False Positive (FP) = 0

False Negative (FN) = 0

Since there are no classification errors, the precision, recall and f1-score are all 1.00 (100%).

Precision = $\frac{\text{TP}}{\text{TP+FP}} = \frac{15}{15+0} = 1.00$

Recall = $\frac{\text{TP}}{\text{TP+FN}} = \frac{15}{15+0} = 1.00$

F1-score = $\frac{2 \times (1.0 \times 1.0)}{1.00+1.00} = 100\%$

The results of the calculation of the Support Vector Machine algorithm prove that the test data is quite accurate and relecant. And the SVM algorithm can achieve 100% accuracy in classifying drug types based on the data used.

### 3.3. Result Analysis

In the background of chapter I, it is explained that improper drug selection can edanger patients, because of the manual process by medical personnel. And often coxtrained by time, information and risk of error. So an artificial intelligence based solution is considered important. Using the Support Vector Machine (SVM) algorithm shows that accuracy, precision, recall and f1-score reach 100%. As well as the SVM algorithm is able to classify drug types quickly and precisely without errors in the test data. Chapter II discusses the theory of the Support Vector Machine (SVM) algorithm, which is able to optimally separate data with the best hyperlane. The results of this research show 100% accuracy, compared to previous research that only achieved 95% accuracy. This research proves that SVM is very effective in classifying drugs against patients. Chapter III describes the method of this research with the KDD (Knowledge Dicovery in Database) stafe starting from data collection, pre-processing, transformation, data mining and evaluation. With a dataset of 200, it is sufficient to train the SVM model. With the KDD stage, it is proven by the success of frug classification for patients.

## 3. Conclusion

The results show that the Support Vector Machine (SVM) algorithm is effective in classifying drug types based on patient medical data. Although the dataset is only 200 data, the SVM model is able to achieve 100% accuracy. Compared to the manual process, the application of SVM proved to be more efficient, consistent and reduced potential errors in drug selection. The use of a linear kernel in the division of training and testing data by 80:20 succeeded in producing a stable and adequate model. And can be as a tool in decision support sistem in the medical field.

## References

[1] S. Febri Aldi, Irohito Nozomi, “Comparison of Drug Type Classification Performance Using KNN Algorithm,” *Sink. (Jurnal Penelit. Tek. Inform.*, 2022, doi: https://doi.org/10.33395/sinkron.v7i3.11487.

[2] F. A. Mardewi Mardewi, Nuru Yarkuran, Sofyan Sofyan, “KLASIFIKASI KATEGORI OBAT MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE,” *J. Pharm. Appl. Comput. Sci.*, 2023, doi: https://doi.org/10.59823/jopacs.v1i1.19.

[3] B. P. S. (BPS), *Profil Statistik Kesehatan 2023*. 2023. [Online]. Available: https://www.bps.go.id/id/publication/2023/12/20/feffe5519c812d560bb131ca/profil-statistik-kesehatan-2023.html

[4] C.-F. H. Hao-Ren Yao, Der-Chen Chang, Ophir Frieder, Wendy Huang, MD, I-Chia Liang, MD, “Cross-Global Attention Graph Kernel Network Prediction of Drug Prescription,” *ACM Digit. Libr.*, 2020, doi: https://doi.org/10.1145/3388440.3412459.

[5] R. K. Avik Sengupta, Saurabh Kumar Singh, “Support Vector Machine–Based Prediction Models for Drug Repurposing and Designing Novel Drugs for Colorectal Cancer,” *ACS Omega*, 2024, doi: https://doi.org/10.1021/acsomega.4c01195.

[6] S. B. Laili Nur Farida, “Klasifikasi Gagal Jantung menggunakan Metode SVM (Support Vector Machine),” *Komputika J. Sist. Komput.*, 2024, doi: https://doi.org/10.34010/komputika.v13i2.11330.

[7] E. B. S. Wisnu Jayanata, Isman Kurniawan, “Implementasi Metode Support Vector Machine (SVM) pada Model Prediksi Rating Obat Berdasarkan Ulasan Pasien,” *Log. J. Penelit. Inform.*, vol. 1, 2023, doi: https://doi.org/10.25124/logic. vlil.6410.

[8] T. H. Nur Ghaniaviyanto Ramadhan, Azka Khoirunnisa, Kurnianingsih, “A Hybrid ROS-SVM Model for Detecting Target Multiple Drug Types,” *JOIV Int. J. Informatics Vis.*, 2023, doi: https://dx.doi.org/10.30630/joiv.7.3.1171.

[9] S. E. M. Purba, “A Comparative Study of Drug Prediction Models using KNN, SVM, and Random Forest,” *ISI J. Inf. Syst. Informatics*, 2025, doi: https://doi.org/10.51519/journalisi.v7i1.1013.

[10] Y. L. Yessy Phalentina br Sagala, Roman Samosir, “Investigation of The Increase in Drug Use in Medan City Using The Support Vector Machine (SVM) Method,” *J. Comput. Networks, Archit. High Perform. Comput.*, 2024, doi: https://doi.org/10.47709/cnahpc.v6i3.4137.

[11] R. H. Dhini Septhya, Kharisma Rahayu, Salsabila Rabbani, Vindi Fitria, Rahmaddeni Rahmaddeni, Yuda Irawan, “Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru,” *Indones. J. Mach. Learn. Comput. Sci.*, 2023, doi: https://doi.org/10.57152/malcom.v3i1.591.

[12] R. H. Mariana Dewi, Triando Hamonangan Saragih, “Penerapan SMOTE-NCL untuk Mengatasi Keisian Kelas pada Klasifikasi Penyakit Jantung Koroner,” *JIP (Jurnal Inform. Polinema)*, 2023, doi: https://doi.org/10.33795/jip.v10i1.1394.

[13] I. A. Rohmatullah Sony Wijaya, Arie Qur'ania, "Klasifikasi Penyakit Cacar Monyet Menggunakan Support Vector Machine (SVM)," *Indones. J. Mach. Learn. Comput. Sci.*, 2024, doi: https://doi.org/10.57152/malcom.v4i4.1417.

[14] E. E. Fienda Altamevia, Harma Oktafia Lingga Wijaya, "Analisis Pola Penjualan Obat di Apotek Srikandi Menggunakan Algoritma Supervised Learning," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, 2023, doi: https://doi.org/10.30645/kesatria.v4i1.127.

[15] S. Tita Risa Dewanti, Rastri Prathivi, "IMPLEMENTASI METODE SVM UNTUK KLASIFIKASI PENYAKIT STUNTING BAYI," *J. Inform. Teknol. dan Sains*, vol. 7, 2025, doi: 10.51401/jinteks.v7i1.5070.

[16] A. D. Dina Elly Yanti, Lizah Framesti, "PERBANDINGAN ALGORITMA C4.5 DAN SVM DALAM KLASIFIKASI PENYAKIT ANEMIA," *JIP (Jurnal Inform. Polinema)*, 2023, doi: https://doi.org/10.33795/jip.v9i4.1381.

[17] A. S. Anita Desiani, Sri Indra Maiyanti, Yuli Andriani, Bambang Suprihatin, Ali Amran, Chika Marselina, "Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor," *J. Process.*, 2023, doi: https://doi.org/10.33998/processor.2023.18.1.700.

[18] Y. S. D. Lutfi Ardining Tyas, I Made Tirta, "PENERAPAN MODEL LEAST SQUARE SUPPORT VECTOR MACHINE (LSSVM) UNTUK PERAMALAN KASUS COVID-19 DI INDONESIA," *J. Gaussian*, 2023, doi: https://doi.org/10.14710/j.gauss.12.2.304-313.

[19] Z. S. H. Hamdalla Fadil Kareem, Muayed S AL-Huseiny, Furat Y. Mohsen, Enam A. Khalil, "Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset," *Indones. J. Electr. Eng. Comput. Sci.*, 2021, doi: http://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738.

[20] S. T. P. Rohit Chivukula, T. Jaya Lakshmi, Sanku Satya Uday, "Classifying clinically actionable genetic mutations using KNN and SVM," *Indones. J. Electr. Eng. Comput. Sci.*, 2021, doi: http://doi.org/10.11591/ijeecs.v24.i3.pp1672-1679.

[21] D. K. Pramudito, M. Tubagus, A. Ismail, F. D. Anggraini, and C. C. Pradnya Paramita, "Application of Support Vector Machine and Naive Bayes Method to Analyze Mobile Banking User Behavior Through social media Twitter.", jidt, vol. 5, no. 4, pp. 325-330, Jan. 2024. doi: https://doi.org/10.60083/jidt.v5i4.462

[22] D. K. Pramudito, S. Arijanti, Z. Akbar, R. Haribowo, and P. Diawati, "The Application of Developed Delone & McLean Method to Analyze The Success and Effectiveness of Alfagift Application Among Big Cities' Users in Indonesia", jidt, vol. 5, no. 4, pp. 111-117, Dec. 2023. doi: https://doi.org/10.60083/jidt.v5i4.424

[23] M. Kasendra, D. K. Pramudito, and D. Afandi, "The Soil Moisture Monitoring Utilize IoT System To Improve Urban Farming Productivity and Enhance Food Security ", jidt, pp. 1-12, Apr. 2025. doi: https://doi.org/10.60083/jidt.vi0.609