TIFDA JOURNAL
Technology Information and Data Analytic
Journal Website : http://jeptira.id/index.php/tifda
e-ISSN : 3064-0660

Implementation of Support Vector Machine and Multilayer Perceptron Algorithms for Patient Diagnosis Based on Patient Profile and Complaints at Jatibening Public Health Center

Muhammad Romanda Ilham[1], Afri Yudha[1]
[1]Information Technology, Faculty of Engineering, University of Darma Persada
[1]Jl. Taman Malaka Selatan No.8 Pd. Kelapa, Kec. Duren Sawit, Kota Jakarta Timur,
Daerah Khusus Ibukota Jakarta, 13450, Indonesia
romandailham@gmail.com

Abstract — Community health centers (Puskesmas) are primary healthcare institutions that play a crucial role in providing services to the community, especially in areas with limited access. However, the disease identification process at the Jatibening Community Health Center still uses traditional methods that are time-consuming and potentially biased. This study aims to create a disease prediction system for patients using the Support Vector Machine (SVM) and Multilayer Perceptron (MLP) machine learning algorithms that utilize data from patient profiles and complaints. The methods used in this study include collecting information from patient medical records, data processing, training SVM and MLP models, and assessing the model's accuracy level. Test results show that the MLP algorithm achieves 100% accuracy, while the SVM also demonstrates 100% accuracy in predicting the likelihood of a patient's disease based on factors such as age, gender, and chief complaint. Thus, the use of machine learning algorithms on patient data at the Jatibening Community Health Center can accelerate the initial diagnosis process and support more efficient medical decision-making.

Keywords: Support Vector Machine, Multilayer Perceptron, Patient Diagnosis, Community Health Center, Medical Records

I. INTRODUCTION

Primary healthcare centers (*Puskesmas*) play a pivotal role in delivering essential health services, especially in areas with limited access to advanced medical facilities [1][2]. Puskesmas increasing number of patients with various health complaints necessitates a diagnostic process that is both fast and accurate to support optimal medical services [3]. Timely and precise diagnosis is crucial, particularly for patients at high risk, in order to provide appropriate treatment and prevent complications [4].

Although patient data is now digitized, its utilization for automated diagnosis remains minimal [5]. Conventional diagnostic methods still rely heavily on manual analysis by medical personnel, which is time-consuming and susceptible to subjectivity [6]. This presents a challenge to the quality and efficiency of healthcare services. Therefore, implementing artificial intelligence (AI) based technology, particularly machine learning algorithms, emerges as a potential solution to improve diagnostic accuracy and speed [7].

Among the various machine learning techniques, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) are widely adopted for medical data classification tasks. SVM excels at creating optimal decision boundaries in high-dimensional spaces and is resistant to overfitting, especially on small datasets [8][9]. MLP, as a form of artificial neural network, is capable of learning complex, nonlinear patterns within medical data, thus offering strong predictive performance in health-related applications [10].

This study aims to implement SVM and MLP algorithms to predict diseases based on patient profiles and complaints at Puskesmas Jatibening. The goal is to support medical decision-making with a system that enhances diagnostic accuracy and reduces the time needed for initial assessments. Ultimately, this research contributes to the development of intelligent decision-support systems that can improve public health service quality.

Alongside the growing adoption of artificial intelligence (AI) in healthcare, several studies have demonstrated its potential to support clinical decisions in resource-limited settings. Applied machine learning for cancer prognosis, while compared the performance of SVM and MLP in health datasets [5]. However, these studies were often conducted in controlled laboratory settings or using large international datasets. In contrast, the present research focuses specifically on a local health center context, where medical staff face time constraints and high patient loads. This context highlights the importance of designing a system that is both technically accurate and practically usable.

II. METHODOLOGY

This study applies a quantitative experimental method to implement and evaluate two machine learning algorithms Support Vector Machine (SVM) and Multilayer Perceptron (MLP) for predicting diseases based on patient profile and complaint data from Puskesmas Jatibening. To structure the research workflow, this study adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology, which consists of six stages:

1. Business Understanding
   The objective is to build a decision-support system that can assist medical staff in diagnosing diseases early and efficiently based on patient-reported data.

2. Data Understanding
   Patient data including age, gender, and main complaints are collected. The data is explored to identify patterns, anomalies, and assess overall structure.

3. Data Preparation
   The collected data is cleaned and processed through normalization, encoding of categorical features, and partitioned into training and testing datasets.

4. Modeling
   The system is developed using the Python programming language. SVM and MLP algorithms are trained using the Scikit-learn library. The input features include patient profile and complaints, while the output is a predicted disease class along with probability confidence.
5. Evaluation
   The trained models are evaluated using standard classification metrics, including accuracy, confusion matrix, and classification report. The performance of SVM and MLP is compared to determine the most suitable algorithm.

6. Deployment
   The final system is deployed as a web-based application using Streamlit and connected to Firebase Realtime Database for real-time storage of prediction results.

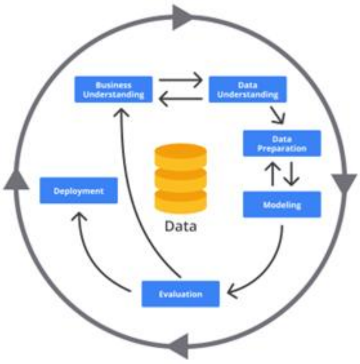
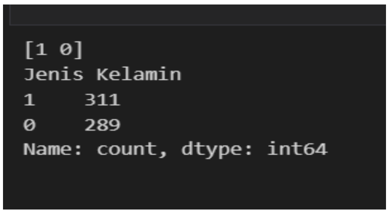


Fig.1. CRISP – DM

The dataset consisted of approximately 600 patient records, with an almost balanced gender distribution (52% female, 48% male). Age ranged from 5 to 70 years, covering both pediatric and adult populations. The most frequent complaints were fever (28%), cough (25%), dizziness (20%), and cold (15%), with other symptoms distributed across the remaining cases. This variety ensured that the model could capture both common and relatively less frequent health complaints.



```
[1 0]
Jenis Kelamin
1    311
0    289
Name: count, dtype: int64
```

Fig.2. Gender in the dataset

III. RESULTS AND DISCUSSION

This section presents the results obtained during the implementation and evaluation stages of the system, following the CRISP-DM methodology. The discussion is divided into seven sub-sections, which cover the complete lifecycle of the system development starting from data acquisition to model training, evaluation, and final deployment. Each stage is elaborated in detail as follows:

A. Data Collection Results

The dataset used in this study was obtained from Puskesmas Jatibening and consists of 600 patient records. Each record includes features such as gender, age, and main complaint. The original data was provided in JSON format and converted into CSV for further preprocessing. After removing incomplete or irrelevant entries, the dataset was prepared for analysis and modeling.

B. Data Analysis Results

Exploratory data analysis revealed that the most frequently reported symptoms were fever, cough, dizziness, and cold. Patients were predominantly within the 5 - 70 years age group, and the distribution of genders was relatively balanced. The complaint field, originally in free-text form, was transformed into structured features using a multi-label encoding strategy, allowing the complaint types to be interpreted as binary features.

C. Data Preprocessing Results

The preprocessing stage involved multiple steps: numerical data (age) was normalized, gender was label-encoded, and complaint data was transformed into binary features using one-hot encoding. The dataset was then split into training and testing sets with an 80:20 ratio, resulting in 600 training and 120 testing records.

D. Model Training Results

Two machine learning models were trained: Support Vector Machine (SVM) and Multilayer Perceptron (MLP). Both models were developed using Scikit-learn in Python. SVM was configured with a linear kernel, while MLP employed three hidden layers and the ReLU activation function. The input features included the encoded patient complaints, age, and gender, and the output was a diagnosis label (class 1, 2, or 3).

E. Model Testing Results

The trained MLP model achieved perfect performance on the testing dataset, with the following metrics.

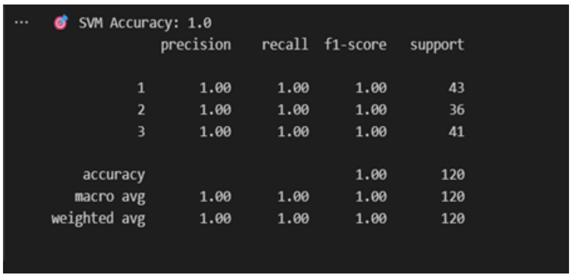
Table 1. Confusion Matrix Table

Diagnosis Class	**Precision**	**Recall**	**F1-score**	**Support**
1	1.00	1.00	1.00	43
2	1.00	1.00	1.00	36
3	1.00	1.00	1.00	41

Overall Accuracy: 100%
Macro Average: 1.00
Weighted Average: 1.00

These results demonstrate the high capability of MLP in recognizing disease patterns from patient data. SVM also produced strong results, but MLP outperformed it in terms of overall accuracy and consistency.

The test result used in this study is the Confusion Matrix. The confusion matrix provides a detailed picture of the predictions made by the model, allowing us to understand the types of errors made and to calculate other important evaluation metrics. This is very useful in comparing the performance of the MLP and SVM models in predicting early lung disease based on processed datashow on Figure 2.



```
...  🎯 SVM Accuracy: 1.0
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        43
           2       1.00      1.00      1.00        36
           3       1.00      1.00      1.00        41

    accuracy                           1.00       120
   macro avg       1.00      1.00      1.00       120
weighted avg       1.00      1.00      1.00       120
```

Fig.3. SVM Model

F. System Implementation Results

The final system was deployed as a web-based application using the Streamlit framework. Users can input patient data through a simple interface, choose between SVM and MLP models, and receive instant prediction results. Each prediction is stored automatically in Firebase Realtime Database**,** enabling real-time logging and future analysis. This implementation makes the system accessible, scalable, and usable for medical staff without technical expertise.

The deployment of the system through Streamlit with Firebase integration ensured that the predictions were immediately stored in the cloud. This functionality enables tracking patient history, conducting further analysis of recurring symptoms, and facilitating communication between medical staff. The cloud-based approach also makes the system scalable, allowing it to be extended to other Puskesmas in the future without significant reconfiguration.

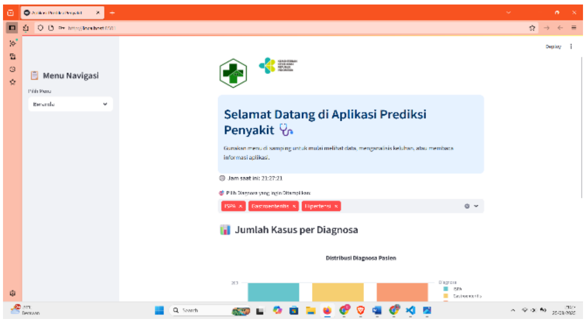
```
# ===============================
# Konfigurasi Firebase
# ===============================
firebaseConfig = {
    "apiKey": "AIzaSyAFVY0GjcLzrg5PrrbmBeZWXJkZGfgihcU",
    "authDomain": "skripsi-86872.firebaseapp.com",
    "databaseURL": "https://skripsi-86872-default-rtdb.firebaseio.com/",
    "projectId": "skripsi-86872",
    "storageBucket": "skripsi-86872.appspot.com",
    "messagingSenderId": "722934043351",
    "appId": "1:722934043351:web:89b5b3794efd96ea519692",
    "measurementId": "G-E1RCLT7E9H"
}
firebase = pyrebase.initialize_app(firebaseConfig)
db = firebase.database()
```

Fig.4. Configuration Firebase and Streamlit

Figure 4 illustrates the main dashboard of the disease prediction application developed using Streamlit. On this homepage, medical staff are welcomed with general information about the system and can immediately select which diagnoses they want to display.

The system also provides a visualization of the number of cases per diagnosis, enabling healthcare workers to easily observe disease distribution trends at the community health center. From the initial testing phase, it can be seen that ISPA, Gastroenteritis, and Hypertension were among the most frequently reported conditions in the patient dataset.

This feature is particularly valuable for medical staff, as it allows them to monitor case distribution in real-time. By presenting diagnosis results in a visual form, doctors and healthcare officers can more quickly identify emerging disease patterns and plan appropriate preventive or curative measures.



Selamat Datang di Aplikasi Prediksi Penyakit

Gunakan menu di samping untuk mulai melihat data, menganalisis keluhan, atau membaca informasi aplikasi.

Jam saat ini: 21:27:21

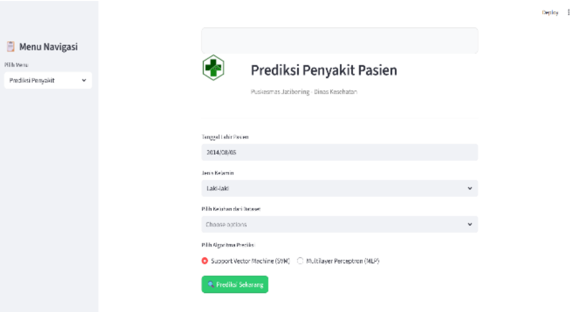
Pilih Diagnosa yang ingin Ditampilkan: ISPA, Gastroenteritis, Hipertensi

Fig.5. Dashboard

The figure below presents the patient disease prediction interface, where users can input relevant patient information such as date of birth, gender, and reported symptoms. This input form ensures that the prediction process is both structured and user-friendly, allowing healthcare staff to quickly enter patient data without technical difficulties.

In addition, the system provides an option to select the prediction algorithm to be used, either Support Vector Machine (SVM) or Multilayer Perceptron (MLP). This feature not only enables flexibility in model usage but also supports comparative evaluation between different machine learning approaches.

Once the required information is filled in, the prediction can be executed by clicking the "Predict Now" button. The output will then provide a diagnosis suggestion, which can support medical staff in making faster and more accurate initial assessments.



Prediksi Penyakit Pasien

Puskesmas Jatibening - Dinas Kesehatan

Tanggal Lahir Pasien: 2014/08/05

Jenis Kelamin: Laki-laki

Pilih Keluhan dari Dataset: Choose options

Pilih Algoritma Prediksi: ● Support Vector Machine (SVM) ○ Multilayer Perceptron (MLP)

Fig.6. Disease Prediction

IV. CONCLUSION

The research successfully demonstrates the implementation of Support Vector Machine (SVM) and Multilayer Perceptron (MLP) algorithms for disease prediction based on patient profile and complaint data from Puskesmas Jatibening. A total of 600 patient records were processed and modeled using the CRISP-DM framework, which ensured a systematic approach throughout the development lifecycle.

Both SVM and MLP models performed well in classifying the data, with the MLP model achieving 100% accuracy, precision, recall, and F1-score on the testing dataset. This indicates that MLP was highly effective in identifying the correct diagnosis from structured input features such as age, gender, and complaints. The results confirm that machine learning can significantly enhance the accuracy and efficiency of initial medical diagnosis in primary healthcare settings.

Additionally, the deployment of the system using Streamlit and Firebase Realtime Database provides an accessible platform for medical staff to utilize prediction results in real time. The system not only assists in decision-making but also facilitates the recording and monitoring of diagnostic outcomes.

In conclusion, machine learning-based diagnosis systems have great potential to improve healthcare services in local health centers. Future work may involve expanding the dataset, incorporating more medical attributes, and conducting validation in real-world clinical settings.

This study highlights that both algorithms can support early disease detection at the community health center level. The implementation of a cloud-based, interactive diagnostic system demonstrates the feasibility of bridging academic machine learning research with real-world healthcare applications. Nevertheless, the findings should be interpreted cautiously due to the limited dataset. Expanding the dataset and validating it in clinical trials with real patients will be essential for wider adoption.

In future work, ensemble learning approaches such as Random Forests, Gradient Boosting, or even hybrid deep learning models could be tested for comparison. Furthermore, incorporating additional patient attributes such as medical history, lifestyle factors, or laboratory results could improve the robustness of the prediction system.

REFERENCES

[1] S. Wahyuni and L. Ferial, "Pemeriksaan Puskesmas di Daerah Terpencil terhadap Fasilitas Kesehatan," *JOURNAL OF BAJA HEALTH SCIENCE*, vol. 3, no. 01, 2023, doi: 10.47080/joubahs.v3i01.2487.

[2] A. R. Luthfia and E. N. S. Alkhajar, "PRAKTIK PELAYANAN PUBLIK: PUSKESMAS SEBAGAI GARDA TERDEPAN PELAYANAN KESEHATAN," *Decision: Jurnal Administrasi Publik*, vol. 1, no. 2, 2019, doi: 10.23969/decision.v1i2.1802.

[3] Amirotul Luthfiyah and Imelda Dian Rahmawati, “Analisis Kualitas Pelayanan Kesehatan di Puskesmas Porong Kabupaten Sidoarjo,” *Jembatan Hukum : Kajian ilmu Hukum, Sosial dan Administrasi Negara*, vol. 1, no. 3, pp. 212–220, Jul. 2024, doi: 10.62383/jembatan.v1i3.533.

[4] R. Gustriansyah, S. Puspasari, A. Sanmorino, N. Suhandi, and F. Antony, “Penerapan Aplikasi Laboratorium Untuk Meningkatkan Kualitas Layanan Di Puskesmas Gandus,” *Jurdimas (Jurnal Pengabdian Kepada Masyarakat) Royal*, vol. 6, no. 3, 2023, doi: 10.33330/jurdimas.v6i3.2425.

[5] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, “Machine learning applications in cancer prognosis and prediction,” 2015. doi: 10.1016/j.csbj.2014.11.005.

[6] M. Z. Nurrokhman, “Perbandingan Algoritma Support Vector Machine dan Neural Network untuk Klasifikasi Penyakit Hati,” *The Indonesian Journal of Computer Science*, vol. 12, no. 4, 2023, doi: 10.33022/ijcs.v12i4.3274.

[7] N. Sharma and H. Om, “Using MLP and SVM for predicting survival rate of oral cancer patients,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 3, no. 1, 2014, doi: 10.1007/s13721-014-0058-x.

[8] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, “Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms,” 2023. doi: 10.3390/electronics12081789.

[9] P. Cortez, “Data Mining with Multilayer Perceptrons and Support Vector Machines,” *Intelligent Systems Reference Library*, vol. 24, 2012, doi: 10.1007/978-3-642-23242-8_2.

[10] N. Uma, “A Hybrid Approach for Medical Image Classification using Support Vector Machines and Artificial Neural Networks,” in *Proceedings of the 4th International Conference on Smart Electronics and Communication, ICOSEC 2023*, 2023. doi: 10.1109/ICOSEC58147.2023.10276063.