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

Comparative Analysis of Naïve Bayes and SVM Algorithms for Early Detection of Lung Disease at Cimuning Community Health Center

Jordi Hafidz Azis[1], Aji Setiawan[2]
[1,2] Department of Information Technology, Faculty of Engineering, Darma Persada University
[1,2] South Malaka Park Road No. 8, East Jakarta 13450, Indonesia
jordihafidz2908@gmail.com

Abstract — This study develops an early prediction model for lung disease using the Naïve Bayes algorithm and Support Vector Machine (SVM) at the Cimuning Health Center to overcome the limitations of diagnostic tools and experts. The evaluation results show that Naïve Bayes is superior with accuracy, precision, recall, and F1-score of 1.00. The research methods include business understanding, data preparation, modeling, evaluation, and implementation, with patient data that has been cleaned and analyzed. The confusion matrix is used to assess model performance. The results show that Naïve Bayes is effective in early prediction of lung disease, so this system is expected to help health centers in detecting patients who need to be referred to the hospital.

Keywords – Naïve Bayes, Machine Learning, Confusion Matrix, Lung Disease, Support Vector Machine.

Copyright © 2025 JURNAL TIFDA
All rights reserved.

I. INTRODUCTION

Lung diseases, such as tuberculosis, bronchitis, and pneumonia, are serious health problems and are one of the main causes of morbidity and mortality in Indonesia. Early detection of lung disease is very important to prevent more severe complications and increase the chances of recovery for patients. However, in primary health facilities, such as the Cimuning Health Center, the diagnosis process is often hampered by limited diagnostic tools, such as X-rays, and a limited number of experts.

Public awareness of the dangers of lung disease and its symptoms is still low. If left without proper treatment, this condition can lead to death. The number of people with lung disease in Indonesia is still high. Based on Globocan or International Agency for Research on Cancer data in 2012, there were 25,322 cases of lung cancer in men and 9,374 cases in women in Indonesia [1].

Along with the development of technology, machine learning has been widely used in the medical field to help the process of diagnosing diseases. The Naïve Bayes algorithm and Support Vector Machine (SVM) are classification methods that are widely used in disease prediction based on patient clinical data. Therefore, this study aims to develop an early prediction model for lung disease using both algorithms to help medical personnel identify patients at risk.

Research in comparing the performance of the Decision Tree algorithm and Support Vector Machine (SVM) in classifying or detecting early lung cancer was conducted by Septhya, et al in 2023 with a focus on finding algorithms that are more accurate and effective in predicting or classifying the risk of lung cancer. However, there are still some limitations in this study including the number and variety of datasets that may not reflect the population at large, so that it can affect the generalization of the model [2].

Research by Haffandi, et al (2023) discussed the classification of lung diseases using the Naïve Bayes Classifier algorithm [3]. This research produces an accurate classification to support early diagnosis, so that medical personnel can make decisions more quickly and accurately in dealing with lung disease and shows that the model achieves an accuracy rate of 97.06%. The limitation of this research lies in the single use of the Naïve Bayes algorithm, while other

methods have the potential to provide different or more optimal results in the classification of lung diseases.

A comparison of decision tree, svm and naive bayes algorithms to predict liver disease was conducted by Herisnan, et al (2022) with the aim of determining which algorithm has the best accuracy and reliability [4]. The methodology used in this study includes a preprocessing stage to clean and prepare data before analysis, including handling missing data and removing duplicates. Feature selection was then applied to filter out variables that have less influence on the prediction, with the aim of reducing the risk of overfitting. After the feature selection process was completed, the data was divided into training and test data with a ratio of 50:50, 70:30, and 90:10 to evaluate the performance of the model in various data sharing scenarios. The three algorithms, namely Decision Tree, SVM, and Naive Bayes, were applied to the processed data. This study has limitations because it only compares three algorithms, while other classification methods can potentially provide better results. In addition, the limited variety of datasets also affects the scope of the findings as well as the applicability of the research results in real-world conditions.

II. METHODOLOGY

This study uses machine learning to develop an early prediction model for lung disease at the Cimuning Health Center, using the Naïve Bayes algorithm and Support Vector Machine (SVM). Figure 1 shows the flow of the research process.

Figure 1. Research process flow

Figure 1 show the methodology flow research includes:

1. Data collection patient medical records with various clinical attributes.
2. Data Preparation, including cleaning, transformation, and data division into training (80%) and testing (20%).
3. Machine Learning Modeling, where Naïve Bayes and SVM are used for model training and optimization.
4. Model Evaluation using accuracy, precision, recall, F1-score, and confusion matrix metrics to assess model performance.
5. Implementation and Analysis of Results, comparing the two algorithms to determine the best model and assessing its effectiveness in helping early diagnosis.
6. Conclusions and Recommendations for further development, such as integration with digital health systems.

Naive Bayes is a classification method that has a high probability of producing optimal system calculations. It relies on the concepts of probability and statistics introduced by British scientist Thomas Bayes. The main principle is to predict future probabilities based on previous experience, known as Bayes' Theorem. The main characteristic of Naïve Bayes Classifier lies in the very strong assumption (naïve) regarding the independence of each condition or event [5].

The Support Vector Machine (SVM) algorithm plays a role in finding the optimal hyperplane, a function that effectively separates two classes. The process involves maximizing the margin or distance between the training pattern and the decision boundary. SVM is not only used for two-class classification, but can also handle cases with a larger number of classes. The One-vs-All approach trains the SVM model for each class by considering it as a positive class, while the other classes are considered negative. In contrast, the One-vs-One approach builds an SVM model for each combination of class pairs. The final prediction result is obtained from the decisions generated by these models [6].

III. RESULTS DISCUSSION

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 Naïve Bayes and SVM models in predicting early lung disease based on processed data show on Figure 2.

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        73
           1       1.00      1.00      1.00        63

    accuracy                           1.00       136
   macro avg       1.00      1.00      1.00       136
weighted avg       1.00      1.00      1.00       136

Confusion Matrix:
[[73  0]
 [ 0 63]]
```

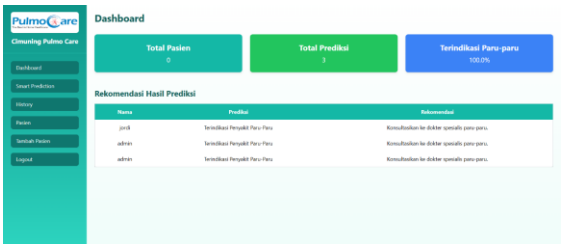
Figure 2. Naïve Bayes Test Results

And there are also test results from the Support Vector Machine (SVM) show on Figure 3.

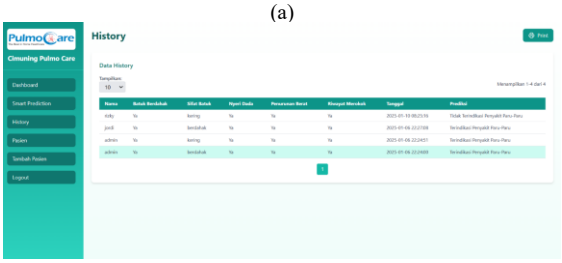
```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.94      0.97        78
           1       0.92      1.00      0.96        58

    accuracy                           0.96       136
   macro avg       0.96      0.97      0.96       136
weighted avg       0.97      0.96      0.96       136

Confusion Matrix:
[[73  5]
 [ 0 58]]
```

Figure3. SVM Test Results

In Figure 4 (a) there is a dashboard view for admin or employee accounts, in this view displays the total registered patients, total predictions, lung indications, and prediction results recommendations. Furthermore, in Figure 4 (b) there is a data history display, where this display shows a data table containing patient names and also the results of input that previously filled in the smart prediction display. and there are also accurate dates and hours when the patient fills in.



(a)



(b)

Figure 4. SVM Test Results (a) & Data History (b)

IV. CONCLUSION

The findings of this research indicate that Naïve Bayes outperforms, allowing for several conclusions to be drawn as follows (1). Comparison of Naïve Bayes and SVM Algorithm Performance: Naïve Bayes outperforms SVM in predicting lung disease early in patients from Cimuning Health Center. Naïve Bayes demonstrates excellent performance with these metrics: testing accuracy 1.00, precision 1.00, recall 1.00, and F1 score 1.00. (2) Elements Influencing Prediction Precision: Aspects like a productive cough, chest discomfort, weight loss, and smoking background must be evenly distributed to accurately forecast early lung illness. This research is in line with research conducted by Naufal, classical machine learning approaches such as Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) can be used to classify diseases [7].

REFERENCES

[1] Wenda, A., Suryanto, A. A., Alam, S. N., & Suhada, K. (2023). Expert System for Diagnosing Lung Diseases Using Bayes Theorem Method. 7, 82–88.

[2] Septhya, D., Rahayu, K., Rabbani, S., & Fitria, V. (2023). Implementation of Decision Tree and Support Vector Machine Algorithms for Lung Cancer Disease Classification. 3(April), 15–19.

[3] Haffandi, M., Haerani, E., Syafria, F., & Oktavia, L. (2022). KLASIFIKASI PENYAKIT PARU-PARU DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER. Jurnal Teknik Informasi Dan Komputer (Tekinkom), 5(2), 176-186. doi:10.37600/tekinkom.v5i2.649

[4] Diva Nabila Herisnan. (2024). KOMPARASI ALGORITMA DECISION TREE, SVM, NAIVE BAYES DALAM PREDIKSI PENYAKIT LIVER. JSR : Jaringan Sistem Informasi Robotik, 8(1), 104–107. https://doi.org/10.58486/jsr.v8i1.38
.

[5] Idris, J. F., Ramadhani, R., Mutoffar, M. M., Tinggi, S., Bandung, T., & Bandung, K. (2024). CLASSIFICATION OF LUNG CANCER DISEASE USING. 2(2).

[6] Setiawan, A., Adi, K., & Catur Edi Widodo. (2023). Rice Variety Identification Using Edge Feature Extraction and Texture-Based Segmentation. 2020 International Seminar on Application for Technology of Information and Communication (ISemantic), 243–249.https://doi.org/10.1109/isemantic59612.2023.10295314.

[7] Naufal, F., Kusuma, S., Tanus, K., Sukiwun. (2021). Analisis Perbandingan Algoritma Klasifikasi Citra Chest X-ray Untuk Deteksi Covid-19 Comparative Analysis of Chest X-ray Image Classification Algorithms for Covid-19 Detection. 10, 96-103.