Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Performance Comparison of Random Forest and SVM in Asthma Prediction

Lailatuz Zuhria[1], Azwar Riza Habibi[2]
[1,2] Department of Informatics Engineering, Asian Institute of Technology and Business, Indonesia
[1]lailazuhria123@email.com, [2] Riza.bj@email.com

Submitted :Dec 15, 2024 | **Accepted** : Jan 1, 2024 | **Published** : Jan 20, 2025

Abstract: This study evaluates the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms in predicting asthma risk to identify the most suitable method for medical datasets. Key metrics include training time, testing time, forecasting time, error rate, and accuracy. The datasets involve attributes such as age and clinical factors, analyzed in three stages: training, testing, and forecasting. Asthma is a chronic respiratory disease affecting quality of life. This study evaluates the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms for predicting asthma risk using clinical and environmental data. The dataset was divided into training, testing, and forecasting subsets. RF showed higher speed in testing and forecasting, while SVM provided lower error rates and higher accuracy. The study concludes that RF is suitable for high-speed applications, while SVM is ideal for precision-critical tasks. These findings contribute to developing decision-support systems in asthma management. During training, SVM demonstrated faster processing, requiring a maximum of 0.5323 seconds compared to RF's 3.7736 seconds on 90% training data. In testing, RF excelled in speed, achieving a minimum testing time of 0.0215 seconds, while SVM required 0.0984 seconds on 10% test data. However, SVM consistently outperformed RF in error rate, with a minimum error rate of 19.17%, compared to RF's 25.10%.

Keywords: Asthma Prediction; Error Rate; Forecasting Time; Medical Datasets; Performance Comparison; Prediction Capabilities; Random Forest; Support Vector Machine; Testing Time; Training Time

INTRODUCTION

Most studies comparing Random Forest (RF) and Support Vector Machine (SVM) algorithms in the context of health data analysis focus on diseases such as diabetes, cancer, or infectious diseases. For example, a study in the IAII (Purbolaksono et al., 2021) Journal demonstrated the use of RF and SVM in predicting diabetes with results showing a high level of accuracy. However, the application of these algorithms in asthma risk detection is still very limited. Previous studies have mostly focused on datasets with linear patterns, such as in diabetes, so they have not explored the performance of RF and SVM in handling asthma datasets that are complex, non-linear, and influenced by environmental and genetic factors. Specific analysis of the effectiveness of RF and SVM for asthma, especially in parameters such as accuracy, error rate, and processing time, has not been a major concern, creating opportunities for further research.

Asthma is a chronic disease of the respiratory tract that has a significant impact on the quality of life of sufferers. It is characterized by primary symptoms such as shortness of breath, coughing, and wheezing that may occur recurrently and in varying degrees of severity. Factors that influence the risk of developing asthma include lifestyle, exposure to allergens or environmental pollutants, and genetic predisposition. As a disease that can affect individuals from different age groups, asthma requires proper management, especially in terms of early diagnosis to prevent serious complications (Gaudillo et al., 2019). Early diagnosis allows for more effective management, including in terms of exacerbation prevention and appropriate therapy planning (Boulet et al., 2019).

Along with technological advances, machine learning-based data analysis has become an important tool in the development of prediction and diagnosis systems in the healthcare field (Rigatti, 2017). One innovative approach in health data analysis is the application of machine learning algorithms to detect the risk of diseases such as asthma. Machine learning offers advantages in analyzing large and complex datasets, thus providing more accurate prediction results than conventional methods (Fadli & Saputra, n.d.). In this context, Random Forest (RF) and Support Vector Machine (SVM) algorithms are the top choices in classification tasks due to their reliability and ability to handle complex data characteristics (Lachaud et al., 2023).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RF is an ensemble-based algorithm that utilizes a combination of multiple decision trees to produce more stable and accurate predictions. It has advantages in handling non-linear data, resistance to overfitting, as well as its ability to manage data containing outliers(Machine Learning: A Brief Introduction to Random Forest, n.d.). Meanwhile, SVM is a margin-based algorithm that focuses on finding the optimal hyperplane to separate data classes (Vishwanathan & Narasimha Murty, n.d.). The main advantage of SVM lies in its ability to handle high-dimensional datasets and provide robust prediction results on data with clear margins (Du et al., 2017). Both algorithms have unique characteristics that make them interesting to compare in the context of disease risk detection.

This study aims to analyze the performance of RF and SVM in detecting asthma disease risk. Researchers utilized a relevant health dataset to compare the two algorithms based on three main parameters, namely accuracy, processing time, and error rate. The analysis was conducted by dividing the dataset into several categories, including training, testing, and forecasting data. With this approach, researchers not only evaluated the performance of each algorithm, but also explored their ability to produce accurate predictions for asthma risk (Khalis Sofi et al., 2021a).

In addition, this research makes an important contribution in selecting a superior algorithm for implementation in medical decision support systems. In some previous studies, RF showed good performance in handling data with high noise, while SVM is often recognized for its effectiveness on datasets that have linear patterns or clear margins. A comparison of these two algorithms is expected to provide greater insight into the strengths and weaknesses of each method, especially in the context of chronic disease risk detection such as asthma.

The results of this study focus on comparing the performance of the Random Forest (RF) and Support Vector Machine (SVM) algorithms in analyzing asthma disease risk data. The authors evaluated both algorithms based on key parameters, such as accuracy, processing time, and error rate, to determine which algorithm is superior in detecting asthma risk (Sanchez-Morillo et al., 2016).

LITERATURE REVIEW

The dataset used in this research comes from Kaggle, with the reference Asthma Disease Dataset, which consists of 2392 data. The initial dataset has 29 attributes, but after going through the preprocessing stage, there are only 10 main attributes that are considered the most relevant for conducting a comparative analysis of the performance of the Random Forest (RF) and Support Vector Machine (SVM) algorithms. The main attributes include Patient ID, which is used to uniquely identify each patient, and Age, which indicates the age of the patient as age can affect the risk level of asthma. In addition, there is Physical Activity, which represents the level of physical activity that affects lung health, and Diet Quality, which reflects the quality of diet and its contribution to overall health. Sleep quality is also an important factor represented by the Sleep Quality variable, which affects the functioning of the patient's immune system. Environmental factors were also taken into account, including Pollution Exposure, Pollen Exposure and Dust Exposure, all of which are potential triggers of respiratory problems. Two variables related to lung function, namely Lung Function FEV1 which measures the volume of air exhaled in one second as an indicator of lung function, and Lung Function FVC which indicates the vital capacity of the lungs by measuring the total volume of air that can be forcibly expelled, were also used.

The preprocessing stage included several steps. The initial data was normalized to reduce skewness by certain relevant methods. Next, blank data is handled using the average imputation method to ensure the integrity of the dataset. After that, the dataset is divided into three parts, namely training data, testing data, and forecasting data by magnitude. The selection of these key attributes was based on clinical relevance and statistical analysis, so that the variables used could provide significant insights for analyzing asthma risk (Du et al., 2017).

METHOD

This research aims to analyze and compare the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms in detecting the risk of asthma disease. The research was conducted through several main stages, namely data preprocessing, modeling using RF and SVM, and performance evaluation based on accuracy parameters, processing time, and Mean Absolute Error (MAE) error rate.

Random Forest (RF) is an ensemble-based algorithm that combines multiple decision trees to improve stability and prediction accuracy (Oshiro et al., n.d.). One important concept in decision trees used by RF is Gini Impurity, which is a measure that evaluates how well a feature separates data by class. Gini Impurity measures the level of "purity" of a node, with lower values indicating that the node is purer (fewer classes mixed in). In the context of RF algorithms, this metric is used to select the best attribute that minimizes impurity during the decision tree formation process. By selecting attributes that reduce impurity, the algorithm can improve the accuracy and robustness of the model in performing classification (Akbar et al., 2020).

Gini Impurity for a node t is calculated using the following formula:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$Gini(t) = 1 - \sum_{i=1}^{C} P_i^2 \quad (1)$$

Where P_i is the probability of class i in a node, and C is the number of classes. The lower the Gini Impurity value, the purer the node is, meaning that there are fewer classes mixed in it. RF algorithms use this measure to build decision trees that reduce impurities in their decision making, ultimately improving the accuracy and robustness of the model in classification.

Table 1 explains what steps occur in the random forest algorithm, and also explains the input and output of the random forest algorithm (Buani, 2024).

Table 1
Algoritma Random Forest

Random Forest Algorithm	
Input	Training dataset: labeled (x, y), where x is the feature vector and y is the target variable n_estimators (Number of trees) Max_depth (Maximum depth of each tree)
Output	Trained random forest model
Step 1	Bootstrap aggregation (bagging): Create multiple subsets of training data by randomly sampling with replacement, so that each subset has the same size as the original training data.
Step 2	Decision tree generation: Build a decision tree for each subset by randomly selecting a subset of features at each node, dividing the nodes based on the best division, and continuing until reaching the maximum depth or minimum number of samples.
Step 3	Prediction: To predict, each decision tree makes a prediction, and then the final result is calculated as the average (regression) or majority vote (classification) of all tree predictions.

Meanwhile, the Support Vector algorithm (SVM) aims to find the optimal hyperplane that maximizes the margin between classes in the data (Pisner & Schnyer, 2020). The decision function for linear SVM is:

$$f(x) = w^T x + b \quad (2)$$

where w is the normal vector to the hyperplane, x is the feature vector, and b is the bias. The margin is maximized when the distance of the closest point of each class to the hyperplane is also maximized. For non-linear SVM, a kernel trick is used to map the data to a higher dimension, allowing separation of data that cannot be separated linearly.

Table 2 explains what steps occur in the random forest algorithm, and also explains the input and output of the random forest algorithm (Du et al., 2017).

Table 2
Support Vector Machine Algorithm

Support Vector Machine Algorithm	
Input	Training dataset: labeled (x, y), where x is the feature vector and y is the target variable Kernel function: a function that maps data points to a higher-dimensional space Regulation parameter (C): Controls the trade-off between maximizing margin and minimizing error
Output	Trained support vector machine model
Step 1	Data transformation: Apply kernel functions to map data to a higher-dimensional space.
Step 2	Find the optimal hyperplane: Find the optimal hyperplane that maximizes the margin between classes, determined by the support vector, which is the closest data to the hyperplane.
Step 3	Prediction: To predict the target variable, calculate the distance of new data points to the optimal hyperplane

The performance evaluation of the two algorithms was conducted using three main methods. Accuracy was calculated as the ratio of the number of correct predictions to the total amount of data (Alves et al., 2019):

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Data} \times 100\% \tag{3}$$

Mean Absolute Error (MAE) is a measure used to evaluate how accurate the prediction model is by comparing the predicted value and the actual value [19]. In the formula as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4}$$

The MAE formula is used to calculate the average of the absolute difference between the actual value (y_i) and the predicted value ($\hat{y}_i$). This absolute difference shows how far the prediction result deviates from the actual value regardless of the direction of deviation, so all deviation values are considered positive. After all the absolute differences are summed up, the result is divided by the number of data (n) to get the average deviation.

The conceptual framework of this research follows a systematic flow, as described in Figure 1 as follows:

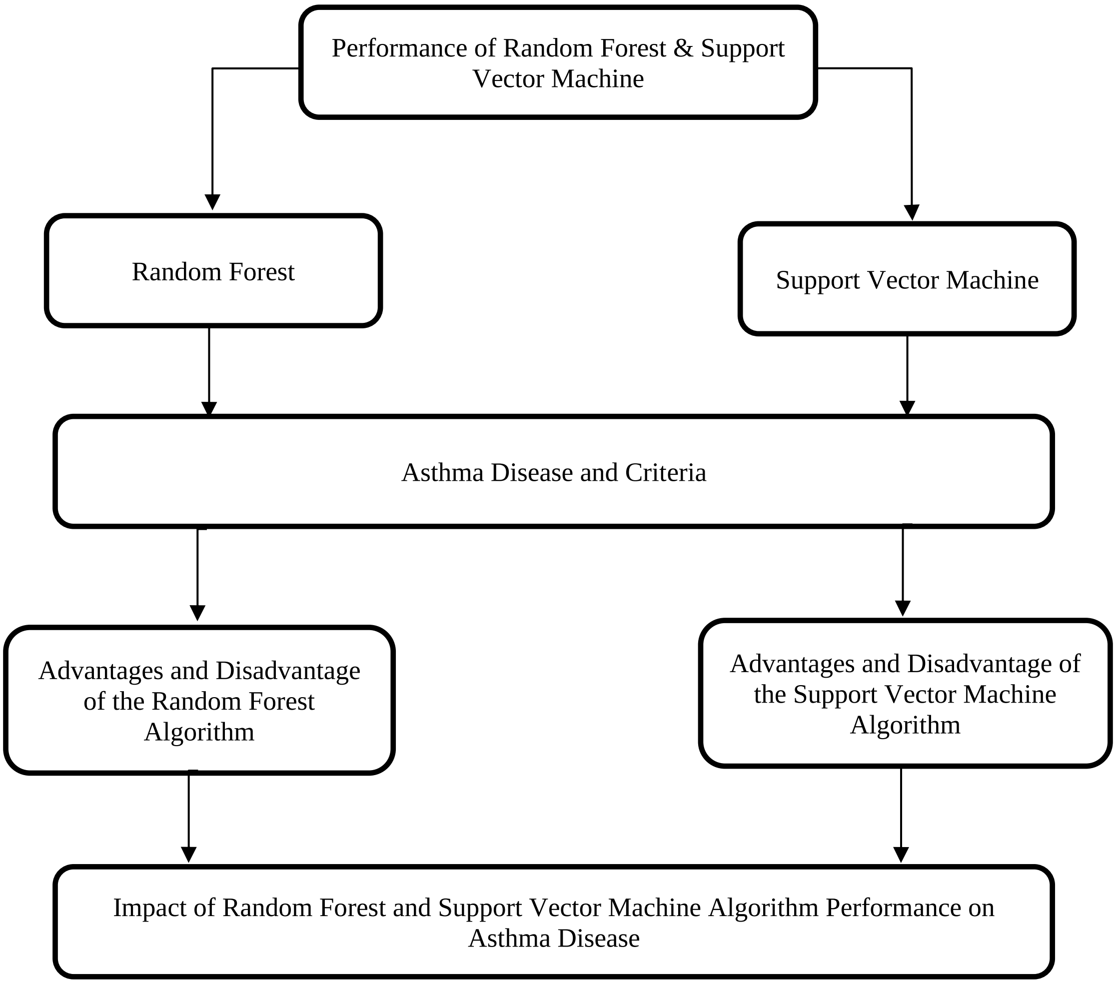

fig. 1 Research Framework

In addition, the research also analyzes the advantages and disadvantages of each algorithm. Random Forest (RF) and Support Vector Machine (SVM) are two popular machine learning algorithms with different characteristics and advantages. Random Forest excels in handling non-linear data, robust to overfitting, and effective in dealing with high-noise data. In addition, RF allows estimation of the importance of features that are

*name of corresponding author

useful in data exploration. However, it has drawbacks such as high computation time on large datasets, lack of result interpretability, and decreased efficiency on high-dimensional datasets. Meanwhile, Support Vector Machine is highly effective for high-dimensional datasets and is able to handle non-linear data through the use of kernel functions. Its main advantage lies in the ability to produce accurate prediction results with maximum margin of separation. However, SVM has limitations in long training time, complex hyperparameter selection, and less efficient performance on large-scale datasets (Khalis Sofi et al., 2021b).

RESULT

In this study, Random Forest (RF) and Support Vector Machine (SVM) algorithms are compared based on their respective performance in three main stages, namely training, testing, and forecasting. Each stage uses a different proportion of data. The analysis includes the time taken by each algorithm to complete each stage as well as the resulting accuracy and error rates. The following is an explanation of the performance of the two algorithms in each stage that has been measured:

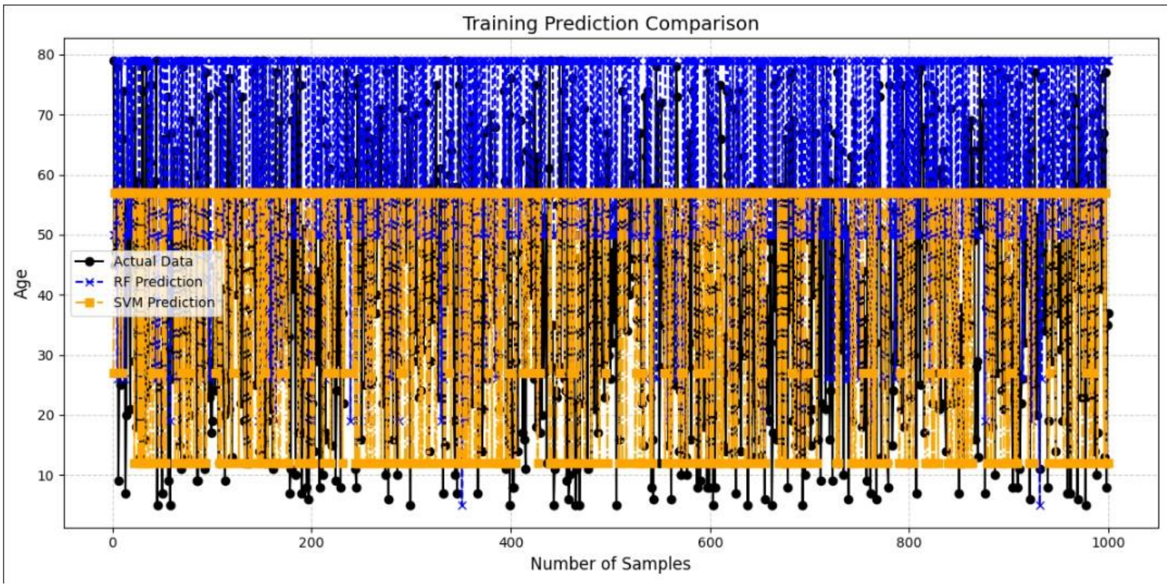

fig. 2 Training Prediction Comparison

Figure 2 explains that this graph shows the comparison of the prediction results of the Random Forest (RF) and Support Vector Machine (SVM) algorithms against the actual data during the training process. The horizontal axis represents the number of samples, while the vertical axis depicts the attribute value (in this case age). The RF prediction pattern appears more volatile and follows the actual data in detail, marked by the crossed blue lines. In contrast, the SVM predictions (orange lines) appear more stable, often staying at a fixed value, indicating that they may be less adaptive to data variations during training.

Table 3 Performance Comparison with Training Data

Algorithm	Training Data Share (Time)				
	60%	70%	75%	80%	90%
Random Forest	2.3650 seconds	3.4311 seconds	2.5571 seconds	3.2315 seconds	3.7736 seconds
SVM	0.3072 seconds	0.8452 seconds	0.4030 seconds	0.4302 seconds	0.5323 seconds

At this stage of the research, various proportions of data were used to compare the algorithms, as shown in Table 3. In the table, the training time is compared between Random Forest (RF) and Support Vector Machine (SVM) algorithms at various proportions of training data (60% to 90%). As a result, SVM is consistently faster than RF across all data shares, with the lowest time of 0.3072 seconds at 60% training data and the highest of 0.5323 seconds at 90% data. In contrast, RF takes longer, ranging from 2.3650 seconds to 3.7736 seconds, due to the nature of the algorithm that builds many decision trees. Overall, SVM is more efficient in terms of training time, but algorithm selection should still consider accuracy and other analysis needs.

*name of corresponding author

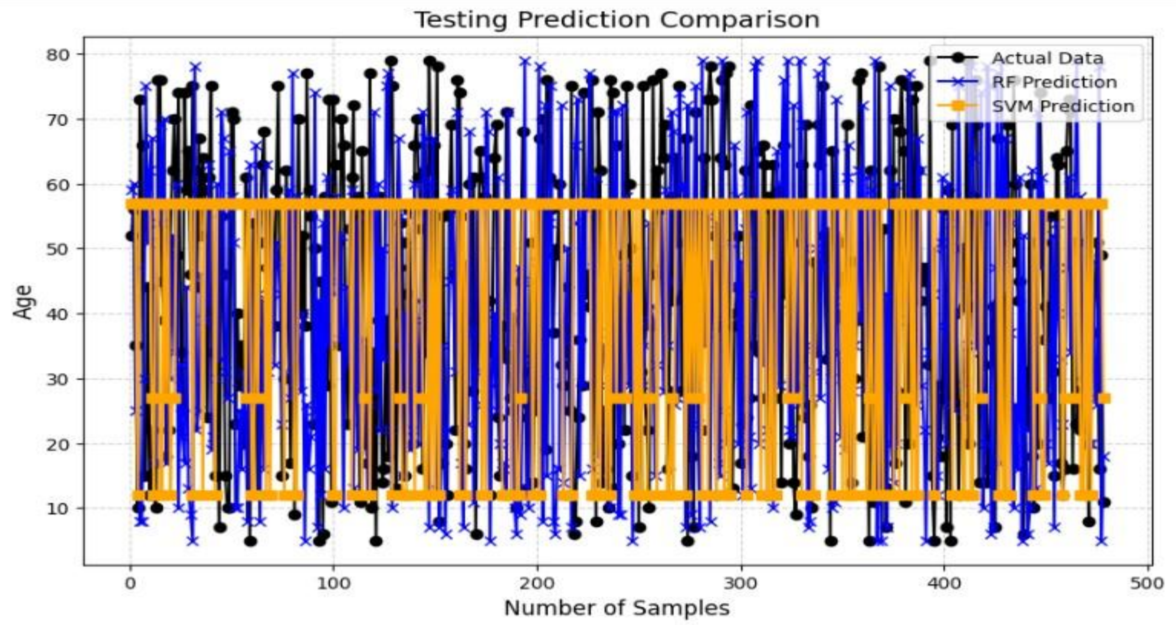

fig. 3 Testing Prediction Comparison

Figure 3 compares the algorithm's prediction performance on forecasting data. RF (crossed blue line) tends to produce predictions that are closer to the actual data (black line), although there are significant fluctuations. SVM (orange line) again shows more stable predictions, but tends not to capture the dynamics of the actual data variations in detail. This may be due to SVM's more conservative approach in generalizing new data.

Table 4. Performance Comparison with Testing Data

Algorithm	Testing Data Share	Testing Time	Testing Accuracy	Testing Error
Random Forest	30%	0.0415 seconds	0.01%	25.7866
SVM		0.1764 seconds	0.01%	19.1227
Random Forest	20%	0.0164 seconds	0.02%	25.1917
SVM		0.0768 seconds	0.01%	24.9417
Random Forest	20%	0.0314 seconds	0.01%	24.4435
SVM		0.1351 seconds	0.01%	19.1715
Random Forest	10%	0.0273 seconds	0.01%	26.159
SVM		0.0740 seconds	0.01%	19.3808
Random Forest	5%	0.0181 seconds	0.03%	24.525
SVM		0.0501 seconds	0.01%	24.9417

In the testing phase, test data was divided into various proportions (30%, 20%, 10%, and 5%), as shown in table 2.4. The table illustrates the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms in testing time, testing accuracy, and testing error. On all divisions of the test data, RF is faster than SVM, for example on 30% of the test data, RF takes 0.0415 seconds, while SVM takes 0.1764 seconds. However, SVM consistently showed lower error rates than RF. For example, on 30% of the test data, the SVM error rate was 19.1227, while the RF reached 25.7866. Overall, RF excels in testing time, but SVM offers better performance in terms of accuracy and error rate.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346

e-ISSN : 2541-2019
p-ISSN : 2541-044X

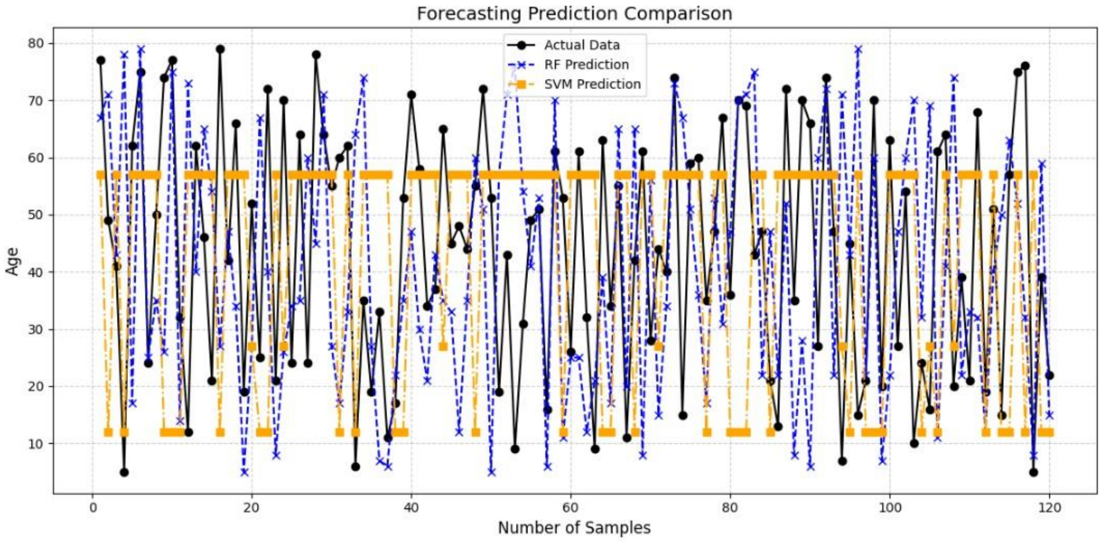

fig. 4 Forecasting Prediction Comparison

And the last stage is the forecasting stage. As shown in figure 4. This graph compares the RF and SVM prediction results against the test data. Just like in the training data, RF provides prediction results that are closer to the actual data, although there are some outliers. SVM still produces a more consistent prediction pattern but is less accurate in capturing the details of value changes in the actual data. The stability of SVM predictions may show advantages in avoiding overfitting, but may come at the expense of accuracy on the test data.

Table 5 Performance Comparison with Forecasting Data

Algorithm	Forecasting Data Share	Forecasting Time	Forecasting Error
Random Forest		0.0268 seconds	23.6333
SVM	10%	0.0618 seconds	19.7708
Random Forest		0.0221 seconds	26.7333
SVM	10%	0.0782 seconds	21.075
Random Forest		0.0195 seconds	23.2583
SVM	5%	0.0377 seconds	19.4917
Random Forest		0.0247 seconds	25.7125
SVM	10%	0.0815 seconds	20.1625
Random Forest		0.0178 seconds	24.1583
SVM	5%	0.0689 seconds	21.075

In the forecasting stage, it is done to predict results based on data that is not included in the training and testing process, as shown in table 2.5. The table shows the performance comparison of Random Forest (RF) and Support Vector Machine (SVM) algorithms in processing forecasting data at 10% and 5% data split. The results show that RF excels in terms of forecasting time speed. For example, at 5% data split, RF takes 0.0195 seconds, while SVM takes 0.0377 seconds. However, SVM has a lower forecasting error than RF. For example, at 10% data split, the SVM error rate is 19.7708, while RF reaches 23.6333. This shows that although RF is faster in calculation, SVM provides more accurate prediction results at the forecasting stage.

DISCUSSIONS

In this research, Random Forest (RF) and Support Vector Machine (SVM) algorithms are applied to detect the risk of asthma disease based on the available dataset. Both algorithms were evaluated by dividing the dataset into subsets for training, testing, and forecasting. RF was used to construct multiple decision trees on subsets of the data and generate predictions by aggregating the results from all trees. Conversely, SVM utilized the hyperplane technique to separate data into different classes, aiming to maximize the margin between classes. This

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

implementation was designed to compare the performance of the two algorithms in terms of processing time, prediction accuracy, and error rate.

The study revealed that Random Forest consistently outperformed SVM in handling datasets with numerous features and high-dimensionality, showing superior prediction stability and lower error rates, especially when the dataset contained noisy or imbalanced data. RF demonstrated better resilience to overfitting and outliers, which contributed to its reliability in prediction. However, the computational time for RF was significantly higher due to the need to build and combine a large number of decision trees. On the other hand, SVM exhibited strong performance on small datasets with complex distributions and clear class separations. Its kernel functions allowed for flexible data mapping into higher dimensions, achieving competitive accuracy in specific scenarios. Nevertheless, SVM was more sensitive to parameter tuning and kernel selection, which could significantly affect its performance. For larger datasets, SVM's computational complexity resulted in prolonged training times, making it less practical compared to RF.

The findings of this study align with previous research on asthma prediction using machine learning. For instance, (Gaudillo et al., 2019) highlighted the robustness of machine learning models like RF in managing genetic and clinical datasets for asthma prediction. Similarly, (Du et al., 2017) demonstrated the efficacy of SVM in handling high-dimensional data, although they noted its computational limitations. This study contributes further evidence to the existing literature by providing a direct performance comparison of RF and SVM in the context of asthma risk detection, reinforcing the suitability of RF for larger, feature-rich datasets and SVM for smaller, well-structured datasets.

Several limitations must be considered in this study. First, the dataset used may not fully represent real-world scenarios, as it was pre-processed and derived from specific sources, potentially limiting the generalizability of the results. Second, the performance of SVM is highly dependent on the choice of kernel function and hyperparameter optimization, which may introduce variability in outcomes if not properly calibrated. Third, Random Forest's computational cost for larger datasets could pose challenges for deployment in resource-constrained environments. Additionally, the study did not include advanced ensemble techniques or hybrid models, which might improve prediction accuracy or reduce computational requirements. Future research could address these limitations by exploring more diverse datasets, optimizing model parameters further, or integrating additional machine learning algorithms for a more comprehensive analysis.

CONCLUSION

The results of this study show a comparison of the performance of the Random Forest (RF) and Support Vector Machine (SVM) algorithms in detecting asthma risk at three stages: training, testing, and forecasting. In the training stage, SVM is more efficient with the lowest time of 0.3072 seconds on 60% of the training data, compared to RF which takes up to 3.7736 seconds on 90% of the data. During testing, RF excels in speed, for example, it takes only 0.0415 seconds on 30% of the test data compared to SVM which takes 0.1764 seconds. However, SVM provided a lower error rate, such as on 30% of the test data, where SVM's error was 19.1227 compared to RF's 25.7866. In the forecasting stage, RF is faster with 0.0195 seconds on 5% of the data than SVM with 0.0377 seconds. However, SVM still has a lower error rate, for example 19.7708 at 10% of the data compared to RF which reaches 23.6333. This study shows that RF is suitable for processing speed, while SVM excels in accuracy. Limitations of the study include the small dataset size and focus on asthma risk. Future studies are recommended to use larger datasets, try other algorithms, or combine RF and SVM for a hybrid approach. This research is useful in improving asthma risk prediction, aiding resource allocation, and improving the quality of data-driven patient care.

REFERENCES

Akbar, W., Wu, W.-P., Faheem, M., Saleem, S., Javed, A., & Saleem, M. A. (2020). Predictive analytics model based on multiclass classification for asthma severity by using Random Forest algorithm. *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1–4. https://doi.org/10.1109/ICECCE49384.2020.9179467

Alves, A. M., Marques de Mello, L., Lima Matos, A. S., & Cruz, Á. A. (2019). Severe asthma: Comparison of different classifications of severity and control. *Respiratory Medicine*, 156, 1–7. https://doi.org/10.1016/j.rmed.2019.07.015

Boulet, L.-P., Reddel, H. K., Bateman, E., Pedersen, S., FitzGerald, J. M., & O'Byrne, P. M. (2019). The Global Initiative for Asthma (GINA): 25 years later. *European Respiratory Journal*, 54(2), 1900598. https://doi.org/10.1183/13993003.00598-2019

Buani, D. C. P. (2024). Deteksi dini penyakit diabetes dengan menggunakan algoritma Random Forest. *Evolusi: Jurnal Sains dan Manajemen*, 12(1). https://doi.org/10.31294/evolusi.v12i1.21005

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14346
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Du, J., Liu, Y., Yu, Y., & Yan, W. (2017). A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms. *Algorithms*, 10(2), 57. https://doi.org/10.3390/a10020057

Fadli, M., & Saputra, R. A. (n.d.). Klasifikasi dan evaluasi performa model Random Forest untuk prediksi stroke. *Jurnal Teknologi*. http://jurnal.umt.ac.id/index.php/jt/index

Gaudillo, J., Rodriguez, J. J. R., Nazareno, A., Baltazar, L. R., Vilela, J., Bulalacao, R., Domingo, M., & Albia, J. (2019). Machine learning approach to single nucleotide polymorphism-based asthma prediction. *PLOS ONE*, 14(12), e0225574. https://doi.org/10.1371/journal.pone.0225574

Khalis Sofi, A. S., Sunge, S. R., Riady, S. R., & Kamalia, A. Z. (2021a). Perbandingan algoritma linear regression, LSTM, dan GRU dalam memprediksi harga saham dengan model time series. *Seminastika*, 3(1), 39–46. https://doi.org/10.47002/seminastika.v3i1.275

Khalis Sofi, A. S., Sunge, S. R., Riady, S. R., & Kamalia, A. Z. (2021b). Perbandingan algoritma linear regression, LSTM, dan GRU dalam memprediksi harga saham dengan model time series. *Seminastika*, 3(1), 39–46. https://doi.org/10.47002/seminastika.v3i1.275

Lachaud, A., Adam, M., & Mišković, I. (2023). Comparative study of Random Forest and Support Vector Machine algorithms in mineral prospectivity mapping with limited training data. *Minerals*, 13(8), 1073. https://doi.org/10.3390/min13081073

Oshiro, T. M., Santoro Perez, P., & Baranauskas, J. A. (n.d.). LNAI 7376 - *How many trees in a Random Forest?* Lecture Notes in Artificial Intelligence. Retrieved from relevant publisher database.

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine Learning* (pp. 101–121). Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 5(2), 393–399. https://doi.org/10.29207/resti.v5i2.3008

Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31–39. https://doi.org/10.17849/insm-47-01-31-39.1

Sanchez-Morillo, D., Fernandez-Granero, M. A., & Leon-Jimenez, A. (2016). Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma. *Chronic Respiratory Disease*, 13(3), 264–283. https://doi.org/10.1177/1479972316642365

Suryanto, A. A. (2019). Penerapan metode mean absolute error (MAE) dalam algoritma regresi linear untuk prediksi produksi padi. *Saintekbu*, 11(1), 78–83. https://doi.org/10.32764/saintekbu.v11i1.298

Vishwanathan, S. V. M., & Narasimha Murty, M. (n.d.). SSVM: A simple SVM algorithm. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02* (Cat. No.02CH37290), 2393–2398. https://doi.org/10.1109/IJCNN.2002.1007516

*name of corresponding author