VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.5272

PERFORMANCE OF ROBUST SUPPORT VECTOR MACHINE CLASSIFICATION MODEL ON BALANCED, IMBALANCED AND OUTLIERS DATASETS

Muhammad Ardiansyah Sembiring[1], Herman Saputra[2], Riki Andri Yusda[3], Sutarman[4*], Erna Budhiarti Nababan[5]

Doctor of Computer Science[1,2,3,4,5]
University of North Sumatra, Indonesia[1,2,3,4,5]
https://fasilkom-ti.usu.ac.id[1,2,3,4,5]
ardiansyah.sembiring@students.usu.ac.id[1], hermansaputra@students.usu.ac.id[2],
riki@students.usu.ac.id[3], sutarman@usu.ac.id[4*], ernabrn@usu.ac.id[5]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *In the realm of machine learning, classification models are important for identifying patterns and grouping data. Support Vector Machine (SVM) and Robust SVM are two types of models that are often used. SVM works by finding an optimal hyperplane to separate data classes, while Robust SVM is designed to deal with uncertainty and noise in the data, making it more resistant to outliers. However, SVM has limitations in dealing with class imbalance and outliers in the dataset. Class imbalance makes the model tend to predict the majority class, and outliers can interfere with model formation. This research compares the performance of SVM and Robust SVM on normal, unbalanced and outlier datasets. The software uses Python and Scikit-learn for implementation and comparison of the two models. Key features include automatic data preprocessing, model training, and evaluation with metrics such as accuracy, precision, recall, and F1 score. The results show that Robust SVM is superior in accuracy on normal datasets and is very effective in dealing with class imbalance, achieving a maximum accuracy of 100%. On datasets with outliers, Robust SVM maintains stable accuracy, demonstrating its robustness to outliers. This research contributes to correspondence management by providing more reliable classification models, improving data processing accuracy, and supporting more informed decision making in software development.*

Keywords*: Machine Learning, Robust SVM, Support Vector Machine.*

Intisari— Di ranah machine learning, model klasifikasi penting untuk mengidentifikasi pola dan mengelompokkan data. Support Vector Machine (SVM) dan Robust SVM adalah dua jenis model yang sering digunakan. SVM bekerja dengan mencari hyperplane optimal untuk memisahkan kelas data, sementara Robust SVM dirancang untuk mengatasi ketidakpastian dan gangguan dalam data, sehingga lebih tahan terhadap outlier. Namun, SVM memiliki keterbatasan dalam menghadapi ketidakseimbangan kelas dan outlier dalam dataset. Ketidakseimbangan kelas membuat model cenderung memprediksi kelas mayoritas, dan outlier dapat mengganggu pembentukan model. Penelitian ini membandingkan kinerja SVM dan Robust SVM pada dataset normal, tidak seimbang, dan dengan outlier. Perangkat lunak menggunakan Python dan Scikit-learn untuk implementasi dan perbandingan kedua model. Fitur utama mencakup prapemrosesan data otomatis, pelatihan model, dan evaluasi dengan metrik seperti akurasi, presisi, recall, dan skor F1. Hasil menunjukkan bahwa Robust SVM lebih unggul dalam akurasi pada dataset normal dan sangat efektif dalam menangani ketidakseimbangan kelas, mencapai akurasi maksimum 100%. Pada dataset dengan outlier, Robust SVM mempertahankan akurasi yang stabil, menunjukkan ketahanannya terhadap pencilan. Penelitian ini berkontribusi pada manajemen korespondensi dengan menyediakan model klasifikasi yang lebih andal,

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5272

meningkatkan akurasi pemrosesan data, dan mendukung pengambilan keputusan yang lebih tepat dalam pengembangan perangkat lunak.

Kata Kunci*: Pembelajaran Mesin, SVM yang Kuat, Mesin Vektor Pendukung.*

INTRODUCTION

In the ever-evolving information age, data processing and analysis are essential for informed decision-making. Classification models, as one of the main branches of machine learning, are important tools in identifying patterns and classifying data into appropriate categories. Two frequently used classification models are Support Vector Machine (SVM) and its specialized variant, Robust SVM. Classification is a technique used to determine or estimate a class or a category of an object based on the attributes or characteristics of the object[1][2]. SVM is a classification algorithm aiming to find the best hyperplane that separates two classes in the feature space. SVM seeks a separator with a maximum margin, which is the greatest distance between the hyperplane and the nearest points from both classes. Robust SVM is a modification of SVM designed to enhance resilience against imbalanced classes and the presence of outliers in data. Outliers, or anomalies, are disruptive data points that can influence the prediction or classification of the data being processed[3][4]. By combining outlier-resistant elements and robust techniques, Robust SVM aims to improve model performance in facing variations and anomalies within a dataset.

In the realm of machine learning, classification models are crucial for identifying patterns and categorizing data into relevant classes. Support Vector Machine (SVM) and its variant, Robust SVM, are widely used classification models. SVM aims to find an optimal hyperplane that separates different classes, while Robust SVM is designed to handle uncertainties and disturbances in the data, making it more resilient to outliers. Class imbalance and the presence of outliers are significant issues in classification tasks. Class imbalance occurs when the number of samples in each class is uneven, often leading the model to favor the majority class and resulting in poor performance on the minority class. Outliers, which are data points that deviate significantly from the rest of the dataset, can distort the model's understanding of the data, leading to inaccurate predictions. Addressing these problems is essential because they can severely degrade the performance and reliability of classification models, hindering their effectiveness in real-world applications.

Despite the widespread use of SVM, its performance can be limited under conditions of class imbalance and outliers. Previous research has demonstrated the effectiveness of SVM in various scenarios, but there is a gap in understanding its performance relative to Robust SVM, particularly in handling these specific issues. This research aims to fill this gap by providing a comparative analysis of SVM and Robust SVM on different types of datasets, including balanced datasets, datasets with imbalanced classes, and datasets with outliers. By doing so, this study provides new insights into the strengths and weaknesses of each model, contributing to the literature on machine learning classification. The primary objective of this research is to compare the performance of SVM and Robust SVM across various dataset conditions. Specifically, this study aims to identify the conditions under which Robust SVM outperforms traditional SVM, providing practical guidance for model selection in scenarios where data quality is compromised by imbalance or outliers. Through this comparison, the research seeks to offer actionable insights that can aid practitioners in choosing the most appropriate model for their specific classification tasks, ultimately enhancing the accuracy and robustness of their machine learning applications.

Previous research[5][6] has investigated the comparison between SVM and Robust SVM. While the results indicate the potential of Robust SVM in enhancing model performance on outlier datasets, a deeper understanding is still needed regarding the extent and under what conditions Robust SVM offers advantages over conventional SVM.

Other earlier studies [7] have reviewed the comparison between SVM and Robust SVM in outlier datasets. Despite promising results, a need for a more profound and comprehensive understanding remains to generate broader and more applicable guidelines for practitioners and researchers. The objective of this research is to compare the performance of SVM and Robust SVM in the context of classification on diverse datasets, including balanced datasets, datasets with imbalanced classes, and datasets with outliers. This study is expected to provide a more comprehensive insight into the performance differences between SVM and Robust SVM in various dataset contexts. The research results can offer practical guidance for selecting the most suitable model based on the characteristics of the faced dataset. Furthermore, this research is anticipated to contribute to the literature on the development of more adaptive and responsive classification models [8]. The objective of this research is to compare the performance of Support Vector Machine (SVM) and Robust SVM in the context of classification across diverse datasets, including balanced datasets, datasets with imbalanced classes, and datasets with outliers. The study aims to provide comprehensive insights into the differences in performance between SVM and Robust SVM models under various dataset contexts, thereby offering practical guidance for selecting the most suitable model based on dataset characteristics. Additionally, this research contributes to the literature by exploring the development of more adaptive and responsive classification models.

MATERIALS AND METHODS

This research method provides a general framework for comparing SVM and Robust SVM in various dataset contexts. The following is the research method used in this study to compare SVM and Robust SVM:

a. Data Collection

In this stage, several datasets covering various characteristics will be selected, including balanced datasets, datasets with imbalanced classes, and datasets with outliers. The datasets are sourced from kaggle.com, and the specific datasets used are the heart dataset for balanced data, the automobile EDA dataset for imbalanced data, and the diabetes dataset for data containing outliers. The selected datasets for this research have different characteristics to test the performance of SVM and Robust SVM models in various dataset contexts. Here are the reasons for selecting the heart, automobile EDA, and diabetes datasets:

Heart Dataset: This dataset is used to represent balanced data cases. As a heart health dataset, it serves as a good representation for situations where the number of samples between positive and negative classes is relatively balanced. Research on this dataset helps in understanding the model performance in situations where the class distribution in the dataset is balanced proportionally.

Automobile EDA Dataset: This dataset is chosen to represent imbalanced data cases. In the real world, we often encounter datasets where one class has a much larger number of samples than the other classes. By using this dataset, the research can test the effectiveness of the model in handling imbalanced class situations.

Diabetes Dataset: This dataset is chosen to represent data cases containing outliers. Outliers are data points that are significantly different from the majority of the data in the dataset. Research on this dataset helps in evaluating the model performance in dealing with disturbances such as outliers, which can affect the accuracy and stability of the model.

By using these three datasets, the research can provide a more comprehensive understanding of the performance of SVM and Robust SVM in various dataset contexts.

b. Data Preprocessing

Subsequently, necessary data preprocessing will be performed, such as handling missing values, normalization, and converting categories into numerical formats.

c. Implementation of SVM and Robust SVM

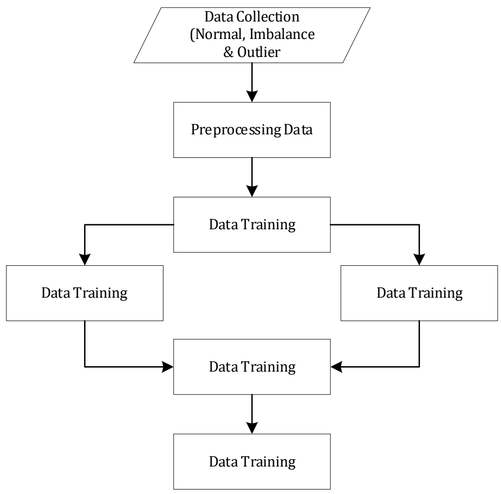
Using Python, in this stage, the implementation of SVM and Robust SVM models will be carried out using available libraries or frameworks, such as Scikit-learn. Both models will then be trained on the training set from each dataset.

d. Model Evaluation

In this stage, the model's performance will be evaluated on the test set using appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score.

e. Comparison Analysis

At this stage, the performance results of SVM and Robust SVM on each dataset will be compared. The analysis aims to identify situations where Robust SVM shows improvement or degradation in performance compared to conventional SVM.



Source: (Research Results, 2024)
Figure 1. Research Flow

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5272

Support Vector Machine

Support Vector Machine is a classification method by maximizes the distance between data classes to find the optimal hyperplane, several kinds of functions of the SVM method are shown in the following equations[9].

a. Linear Kernel

$$K(w_1, w) = w_i^T w \quad (1)$$

b. Gaussian Radial Basic Function Kernel

$$K(w_a, w_b) = exp\left(\frac{\|w_a - w_b\|^2}{2x^2}\right) \quad (2)$$

c. Polynomial Kernel

$$K(w_a, w_b) = exp\big((w_a . w_b) + x\big)^c \quad (3)$$

d. Sigmoid Kernel

$$K(w_a, w_b) = tanh\ \left(\gamma w_i^T w + s\right) \quad (4)$$

The SVM model technique searches for the best hyperplane with the maximum distance to separate two classes, aiming to resolve optimization constraints [10][11]. The decision-making function appears as the following equation.

$$f(x) = z^T m + d = \sum_i \delta_i k\left(m_{i,} m_j\right) + d \quad (5)$$

Where $k\left(m_{i,} m_j\right)$is the kernel function, which measures the similarity or distance between two vectors, variables are Lagrange multipliers, and p is the regularization parameter.

To classify the data, the following equation is utilized [12].

$$f(x_d) = sign\big(g(x)\big) \quad (6)$$

In reality, datasets don't always need to be separable linearly. Data is transformed into linearly separable data through a kernel function and then translated from a narrow space to a feature space.

Robust Support Vector Machine

Support Vector Machine is a classification and regression technique that combines computational algorithms with theoretical foundations[2]. Typically, classification and regression techniques rely on accuracy and efficiency in handling or processing data.

To examine the robustness of the classification method in SVM, the original data is disturbed, including label noise and outliers. Label noise can arise from real-world situations (intentional or unintentional) within individual classes by adding outliers to the dataset [5], with the equation as follows.

$$y_i\left(\sum_{k=1}^{d}(m_k^* x_{ik}) + b^*\right) \quad (7)$$

Based on previous research conducted by Yaqoob[13], a robust scheme for classification has been developed, and proven to be more effective for prediction. The main idea is to construct a robust classification with a higher probability of accuracy for each class. The resulting equation is as follows.

$$\min_{w,b} \frac{1}{2} \|w\|_2^2 \quad (8)$$

Confusion Matrix Measurement

Confusion matrix is a practical method used to assess the performance of a data classification model with known true values. This method is relatively straightforward, but the associated terminology can be confusing[14]. The confusion matrix provides a useful and comprehensive presentation of classifier performance. It is commonly used for multi-class evaluation in single-label classification models, where each data point belongs to only one class at a specific point in time[15]

The concept of organizing a confusion matrix is divided into four parts related to the decisions made along the prediction path as follows[16].

a. True positives (TPH) are the number of points correctly identified as positive by the predicted path.
b. True negatives (TNH) are the number of points correctly identified as negative by the predicted path.
c. False positives (FPH) are the number of points incorrectly identified as positive by the predicted path.
d. False negatives (FNH) are the number of points incorrectly identified as negative by the predicted path.

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i1.5272

The confusion matrix is used to calculate various performance metrics in measuring the effectiveness of a created model [17]. Commonly used performance metrics include accuracy, precision, and recall[18].

Accuracy describes the precision of a model in correctly classifying. It is the ratio of correct predictions (both positive and negative) to the overall data. It can also be said that accuracy is a measure of how close the predictions are to the actual values. The accuracy value is calculated using the following equation.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

Precision is a description of the level of accuracy between the requested data and the results of the predictions made by the model. Precision is the ratio of true positive predictions compared to the overall positive predicted results. The precision value can be obtained through the following equation.

$$precision = \frac{TP}{TP+FN} \quad (10)$$

Recall is a description of the success of the model in retrieving information. Recall can be said to be the ratio of true positive predictions compared to the overall true positive data. The recall value can be obtained through the following equation.

$$recall = \frac{TP}{TP+FN} \quad (11)$$

RESULTS AND DISCUSSION

In this discussion, the analysis process is conducted by initially dividing the data into training and testing sets. The data is split into training and testing sets in an 80:20 ratio. The entire dataset used is selected based on having two labels, namely 0 and 1.

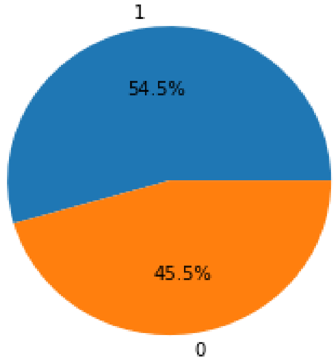
Table 1. Dataset Split

Dataset	Training	Testing	Total
Heart	242	61	303
Automobile EDA	160	41	201
Diabetes	614	154	768

Source: (Research Results, 2024)

a. Performance Comparison of SVM and Robust SVM on Balanced Dataset

The Heart dataset consists of 13 features and 1 label. Before conducting training, it is necessary to visualize the data to observe the percentage for each class in the Heart dataset label.



Source: (Research Results, 2024)
Figure 2: Percentage of Each Class on the Dataset

From Figure 2, the percentage of data for class 0 is depicted as 45.5 percent, and for class 1, it is 54.5 percent. This dataset is considered balanced because it has a relatively small difference in the number of classes. After implementing the SVM and Robust SVM models, a comparison result is obtained based on the classification report from the confusion matrix as follows.

	precision	recall	f1-score	support
0	0.79	0.52	0.62	29
1	0.67	0.88	0.76	32
accuracy			0.70	61
macro avg	0.73	0.70	0.69	61
weighted avg	0.73	0.70	0.69	61

Source: (Research Results, 2024)
Figure 3. Classification Report of SVM Model

	precision	recall	f1-score	support
0	0.86	0.83	0.84	29
1	0.85	0.88	0.86	32
accuracy			0.85	61
macro avg	0.85	0.85	0.85	61
weighted avg	0.85	0.85	0.85	61

Source: (Research Results, 2024)
Figure 4. Classification Report of Robust SVM model

Based on Figures 3 and 4, it is found that the accuracy of SVM is 0.70 (70%), while the accuracy of Robust SVM is 0.85 (85%). In addition to

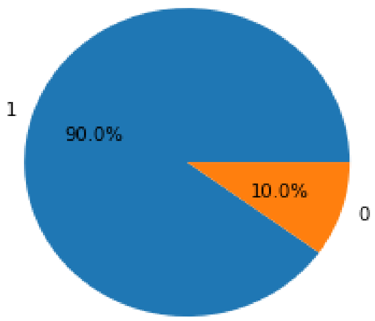
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5272

accuracy, other metrics such as precision, recall, and f1-score in Robust SVM appear to be larger or better. Therefore, it can be concluded that for balanced data both SVM and Robust SVM Performed well, achieving high accuracy. However, robust SVM shower a slight improvement in accuracy compared to traditional SVM. This indicates that while both models are effective in handling balanced data, the robustness of the Robust SVM offers a marginal benefit, potentially due to its ability to better handle slight variations in the data.

b. Performance Comparison of SVM and Robust SVM on Imbalanced Dataset

The automobileEDA dataset consists of 28 features and 1 label. Before conducting training, it is necessary to visualize the data to observe the percentage for each class in the automobileEDA dataset label.



Source: (Research Results, 2024)
Figure 5. Percentage of Each Class on the AutomobileEDA Dataset

From Figure 5, the percentage of data for class 0 is depicted as 10.0 percent, and for class 1, it is 90.0 percent. This dataset is considered imbalanced because it has a significant difference in the number of classes. After implementing the SVM and Robust SVM models, a comparison result is obtained based on the classification report from the confusion matrix as follows.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	6
1	0.85	1.00	0.92	35
accuracy			0.85	41
macro avg	0.43	0.50	0.46	41
weighted avg	0.73	0.85	0.79	41

Source: (Research Results, 2024)
Figure 6. Classification Report of SVM Model

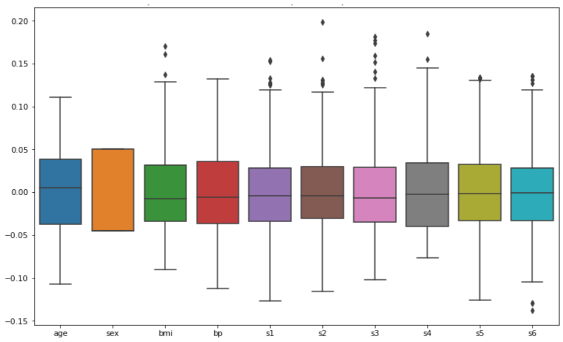
	precision	recall	f1-score	support
0	1.00	1.00	1.00	6
1	1.00	1.00	1.00	35
accuracy			1.00	41
macro avg	1.00	1.00	1.00	41
weighted avg	1.00	1.00	1.00	41

Source: (Research Results, 2024)
Figure 7. Classification Report of Robust SVM Model

Based on figures 6 and 7, it is found that the accuracy of SVM is 0.85 (85%), while the accuracy of Robust SVM is 1.0 (100%). In addition to accuracy, other metrics such as precision, recall, and f1-score in Robust SVM appear to be larger or better. Therefore, it can be concluded that for imbalanced data, the Robust SVM model has better performance results compared to SVM. The performance of SVM significantly dropped on imbalanced datasets, demonstrating its limitation in dealing with class imbalance. The model tended to favor the majority class, leading to poor performance on the minority class. In contrast, Robust SVM maintained higher accuracy and better performance metrics overall. This can be explained by Robust SVM's enhanced capability to adjust to imbalanced data distributions, reducing the bias towards the majority class. The use of robust optimization techniques helps to mitigate the impact of class imbalance by focusing on the misclassified samples and adjusting the decision boundary accordingly.

c. Performance Comparison of SVM and Robust SVM on Outlier Dataset

The diabetes dataset has 8 features and 1 label. Before proceeding with the training, it is necessary to undergo a data visualization process to observe the percentage for each class in the diabetes dataset label.



Source: (Research Results, 2024)
Figure 8. Boxplot for outlier detection in each feature (diabetes dataset).

From Figure 8, it is depicted that several features in the diabetes dataset have outliers. The features with outliers include bmi, s1, s2, s3, s4, s5, and s6. After detecting that the diabetes dataset contains outliers, the next step is to implement the SVM and Robust SVM models, and a comparison result is obtained based on the classification report from the confusion matrix as follows.

	precision	recall	f1-score	support
0	0.77	0.83	0.80	99
1	0.65	0.56	0.60	55
accuracy			0.73	154
macro avg	0.71	0.70	0.70	154
weighted avg	0.73	0.73	0.73	154

Source: (Research Results, 2024)
Figure 9. Classification Report of SVM Model

	precision	recall	f1-score	support
0	0.78	0.84	0.81	99
1	0.66	0.56	0.61	55
accuracy			0.74	154
macro avg	0.72	0.70	0.71	154
weighted avg	0.73	0.74	0.74	154

Source: (Research Results, 2024)
Figure 10. Classification Report of Robust SVM Model

Based on Figures 9 and 10, it is found that the accuracy of SVM is 0.73 (73%), while the accuracy of Robust SVM is 0.74 (74%). In addition to accuracy, other metrics such as precision, recall, and f1-score in Robust SVM appear to be larger or better. Therefore, it can be concluded that for data with outliers, the Robust SVM model has better performance results compared to SVM.

Based on the three comparison processes of SVM and Robust SVM models across three types of data, the accuracy comparison can be summarized in Table 2.

Table 2. Accuracy Summary

Model	Dataset Balance	Dataset *Imbalance*	Dataset *Outlier*
SVM	0.70	0.85	0.73
Robust SVM	0.85	1	0.74

Source: (Research Results, 2024)

By examining the accuracy information in Table 2, both the SVM and Robust SVM models under three different dataset conditions yield several conclusions:

1. Balanced Dataset: The Robust SVM model shows a significant improvement in accuracy for the Balanced Dataset. This indicates that the robust approach can provide tangible benefits in datasets without class imbalance or outliers.
2. Imbalanced Dataset: The Robust SVM model achieves maximum accuracy (1.00) for the Imbalanced Dataset, demonstrating its effectiveness in handling imbalanced classes. The choice of a robust approach has a positive impact on the results.
3. Outlier Dataset: In the Outlier Dataset, the Robust SVM model exhibits good resilience to outliers, maintaining relatively stable accuracy. While the improvement may not be significant, it indicates the robust approach's ability to minimize the impact of outliers.

Overall, these results suggest that employing a robust approach to SVM models can enhance performance, especially in situations involving imbalanced classes and the presence of outliers. In most cases, considering these factors and using a robustly optimized model can yield better outcomes. In this study, a linear kernel is used because the data is separated linearly to achieve optimal results.

CONCLUSION

Based on the conducted research comparing SVM and Robust SVM classification models, it can be concluded that the Robust SVM model outperforms the SVM model across balanced, imbalanced, and outlier datasets. This conclusion is drawn from the testing results using the confusion matrix, which depicts accuracy, precision, recall, and F1-Score, all indicating that the performance of Robust SVM is more optimal than SVM. However, further studies, such as using larger datasets, can be carried out by researchers to gain a more comprehensive understanding of evaluating the Robust SVM model in classification tasks.

REFERENCE

[1] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, “Selecting critical features for data classification based on machine learning methods,” *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00327-4.

[2] N. I. Nordin *et al.*, “Enhancing COVID-19 Classification Accuracy with a Hybrid SVM-LR Model,” *Bioengineering*, vol. 10, no. 11, pp. 1–15, 2023, doi: 10.3390/bioengineering10111318.

VOL. 10. NO. 1 AUGUST 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i1.5272

[3] V. Blanco, A. Japón, and J. Puerto, "A mathematical programming approach to SVM-based classification with label noise," *Comput. Ind. Eng.*, vol. 172, no. May, 2022, doi: 10.1016/j.cie.2022.108611.

[4] A. Gaddam, T. Wilkin, M. Angelova, and J. Gaddam, "Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions," *Electron.*, vol. 9, no. 3, pp. 1–15, 2020, doi: 10.3390/electronics9030511.

[5] M. Baldomero-Naranjo, L. I. Martínez-Merino, and A. M. Rodríguez-Chía, "A robust SVM-based approach with feature selection and outliers detection for classification problems," *Expert Syst. Appl.*, vol. 178, p. 115017, 2021, doi: 10.1016/J.ESWA.2021.115017.

[6] D. Faccini, F. Maggioni, and F. A. Potra, "Robust and Distributionally Robust Optimization Models for Linear Support Vector Machine," *Comput. Oper. Res.*, vol. 147, no. December 2021, p. 105930, 2022, doi: 10.1016/j.cor.2022.105930.

[7] M. Safaei *et al.*, "A systematic literature review on outlier detection inwireless sensor networks," *Symmetry (Basel).*, vol. 12, no. 3, pp. 1–40, 2020, doi: 10.3390/sym12030328.

[8] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J. Pet. Sci. Eng.*, vol. 200, no. December 2020, p. 108182, 2021, doi: 10.1016/j.petrol.2020.108182.

[9] H. Hairani, "Peningkatan Kinerja Metode SVM Menggunakan Metode KNN Imputasi dan K-Means-Smote untuk Klasifikasi Kelulusan Mahasiswa Universitas Bumigora," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 713–718, 2021, doi: 10.25126/jtiik.2021843428.

[10] M. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, "Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset," *Comput. Methods Programs Biomed. Updat.*, vol. 4, no. August, p. 100118, 2023, doi: 10.1016/j.cmpbup.2023.100118.

[11] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. December 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.

[12] A. T. Akbar, N. Yudistira, and A. Ridok, "Identifikasi Gagal Ginjal Kronis dengan Mengimplementasikan Metode Support Vector Machine beserta K-Nearest Neighbour (SVM-KNN)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 301–308, 2023, doi: 10.25126/jtiik.20231026059.

[13] A. Yaqoob, R. M. Aziz, N. K. Verma, P. Lalwani, A. Makrariya, and P. Kumar, "A Review on Nature-Inspired Algorithms for Cancer Disease Prediction and Classification," *Mathematics*, vol. 11, no. 5, 2023, doi: 10.3390/math11051081.

[14] N. E. Ramli, Z. R. Yahya, and N. A. Said, "Confusion Matrix as Performance Measure for Corner Detectors," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 29, no. 1, pp. 256–265, 2022, doi: 10.37934/araset.29.1.256265.

[15] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," pp. 01–14, 2020, doi: 10.5121/csit.2020.100801.

[16] K. Riehl, M. Neunteufel, and M. Hemberg, "Hierarchical confusion matrix for classification performance evaluation," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 72, no. 5, pp. 1394–1412, 2023, doi: 10.1093/jrsssc/qlad057.

[17] T. S. Tamir *et al.*, "Traffic Congestion Prediction using Decision Tree, Logistic Regression and Neural Networks," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 512–517, 2020, doi: 10.1016/j.ifacol.2021.04.138.

[18] K. S. Nugroho, A. Y. Sukmadewa, A. Vidianto, and W. F. Mahmudy, "Effective predictive modelling for coronary artery diseases using support vector machine," *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, pp. 345–355, 2022, doi: 10.11591/ijai.v11.i1.pp345-355.