JURNAL SISFOTENIKA
Vol. 15 No. 1, Januari 2025
p-ISSN: 2087-7897 ; e-ISSN : 2460-5344
DOI : 10.30700/sisfotenika.v15i1.532

Prediksi ISPU Berbasis SVM untuk Masa Depan Jakarta yang Berkelanjutan

SVM-Driven ISPU Prediction for Jakarta's Sustainable Future

Aditya Pratama*[1], Candra Gudiato[2], Adha Maulana[3], Wahyu Prayitno[4]
[1,3]Department of Information System, West Kalimantan Nahdlatul Ulama University, Kubu Raya
[2]Department of Information Technology, Shanti Bhuana Institute, Bengkayang
[4]Department of Environmental Engineering, West Kalimantan Nahdlatul Ulama University, Kubu Raya
e-mail: ***[1]adityapratamabadra@unukalbar.ac.id**, [2]candra.gudiato@shantibhuana.ac.id, [3]adhamaulana@unukalbar.ac.id, [4]wahyuprayitno@unukalbar.ac.id

Abstrak

Polusi udara di Jakarta menimbulkan risiko kesehatan yang signifikan, dengan transportasi sebagai kontributor utama. Studi ini menangani kebutuhan mendesak akan peramalan kualitas udara yang akurat dengan mengembangkan model Support Vector Machine (SVM) untuk memprediksi kategori Indeks Standar Pencemaran Udara (ISPU). Menggunakan data kualitas udara yang luas, termasuk konsentrasi PM10, PM2.5, SO2, CO, dan O3, metode feature selection dan optimasi hiperparameter yang diterapkan pada penelitian ini. Model SVM sangat akurat dalam memprediksi kategori ISPU dengan akurasi (93,84%), dan sangat baik dalam mengklasifikasikan ke dalam kategori "SANGAT TIDAK SEHAT", yang mendapatkan skor sempurna untuk presisi, recall, dan F1-score (1,00). Model ini juga berkinerja baik dalam memprediksi kategori 'SEDANG' dan 'TIDAK SEHAT', yang semakin menunjukkan keandalannya dan presisinya. Hasil dari metode cross-validation menegaskan kekokohan model tersebut. Metodologi ini secara signifikan mengungguli teknik yang ada, menawarkan alat yang efektif untuk manajemen kualitas udara perkotaan. Kemampuan model untuk memprediksi tingkat polusi dengan akurat memungkinkan intervensi kesehatan masyarakat dan strategi perencanaan kota yang lebih terarah. Penelitian ini berkontribusi pada bidang yang berkembang dari aplikasi pembelajaran mesin dalam pemantauan lingkungan dan menyediakan kerangka kerja yang dapat direplikasi untuk megakota lain yang menghadapi tantangan kualitas udara serupa. Pekerjaan mendatang akan fokus pada integrasi data real-time dan eksplorasi metode ensemble untuk meningkatkan akurasi prediktif, sehingga memperbaiki sistem manajemen kualitas udara dan standar kesehatan di Jakarta.

Kata kunci—Support Vector Machine (SVM), Indeks Standar Pencemaran Udara (ISPU), Manajemen Kualitas Udara Perkotaan, Pembelajaran Mesin, Pemantauan Lingkungan

Abstract

Air pollution in Jakarta presents significant health risks, with transportation being a major contributor. This study addresses the critical need for accurate air quality forecasting by developing a Support Vector Machine (SVM) model to predict Air Pollution Standard Index (ISPU) categories. Utilizing extensive air quality data, including PM10, PM2.5, SO2, CO, and

Copyright © 2025 SISFOTENIKA

O3 concentrations, sophisticated feature selection methods and hyperparameter optimization were employed. The SVM model was amazingly accurate at predicting ISPU categories (93.84%), and it did especially well at putting things into the "SANGAT TIDAK SEHAT" category, which got perfect scores for precision, recall, and F1-score (1.00). This model also performed well in predicting 'SEDANG' and 'TIDAK SEHAT' categories, further demonstrating its reliability and precision. The results from cross-validation methods reaffirmed the robustness of the model. This methodology significantly outperforms existing techniques, offering an effective tool for urban air quality management. The model's ability to accurately forecast pollution levels allows for more targeted public health interventions and urban planning strategies. This research contributes to the growing field of machine learning applications in environmental monitoring and provides a replicable framework for other megacities facing similar air quality challenges. Future work will focus on integrating real-time data and exploring ensemble methods to enhance predictive accuracy, thereby improving the air quality management systems and health standards in Jakarta.

Keywords—Support Vector Machine (SVM), Air Pollution Standard Index (ISPU), Urban Air Quality Management, Machine Learning, Environmental Monitoring

1. INTRODUCTION

Poor air quality is one of the biggest challenges faced by major cities around the world, including Jakarta. With the continuously increasing population and rapid urbanization, air pollution in Jakarta has reached alarming levels that threaten public health and the environment. According to a report from the DKI Jakarta Environmental Agency, the transportation sector contributes 67.04% of the total air pollution in Jakarta, indicating the need for effective interventions to manage the continuously declining air quality. Unmanaged emissions from motor vehicles, industries, and construction activities often cause air pollution in Jakarta.

The Air Pollution Standard Index (ISPU) serves as an important tool to inform the public about pollution levels and related health risks. ISPU provides clear categories of air quality, such as good, moderate, unhealthy, and very unhealthy, which help people take preventive measures. By having accurate information about air quality, the community can plan their daily activities, especially for vulnerable groups such as children, the elderly, and individuals with respiratory issues. However, there are still challenges in accurately predicting ISPU categories, especially in dynamic urban contexts like Jakarta.

Accurate air quality predictions are crucial. An uncertainty in air quality data can have a negative impact on public health and affect the correct decision-making by the government and stakeholders. Therefore, the development of predictive models that provide accurate and real-time information is very important. In this context, machine learning technology offers a promising solution for modeling and predicting air quality patterns.

In this study, we propose the use of the Support Vector Machine (SVM) algorithm to predict ISPU categories in DKI Jakarta. SVM is known as an effective method for handling non-linear data and has excellent generalization capabilities, making it an appropriate choice for air quality analysis. SVM works by finding the optimal hyperplane to separate the data into different groups, and it is very suitable for classification problems. In this context, we will use historical data on pollutants such as PM10, PM2.5, SO2, CO, and O3 to create an accurate and reliable predictive model regarding this matter.

This research aims to evaluate the effectiveness of SVM in predicting ISPU categories in Jakarta, as well as to compare it with other machine learning methods. We hope to achieve a higher level of accuracy and precision than traditional approaches by using comprehensive data and advanced feature selection techniques. Additionally, we will optimize the model parameters to enhance prediction performance.

Through this study, we hope to make a significant contribution to air quality management in Jakarta and provide useful tools for policymakers and the general public in facing the increasingly challenging issue of air pollution. We expect the findings of this study to aid in the formulation of more effective air pollution control policies and to increase public awareness of the importance of maintaining air quality.

In addition, this research also aims to enhance knowledge about the application of machine learning techniques in environmental monitoring and explore the potential of SVM to improve urban air quality predictions. With an accurate and reliable model, we hope these results can encourage the implementation of more advanced and responsive air quality monitoring systems that not only provide information to the public but also support efforts to reduce air pollution more effectively.

In the end, we anticipate that this research will lay a solid foundation for future research in air quality and machine learning, as well as open the door for interdisciplinary studies that can use data and technology to enhance environmental and community health.

2. RELATED WORKS

SVM has proven to be an effective method in predicting air quality, especially in the context of complex and non-linear data. Research shows that SVM can provide accurate and reliable predictions in modeling air quality using historical datasets that include various parameters such as particles, ozone, nitrogen dioxide, and carbon monoxide. [1][2][3][4][5].

The advantages of Support Vector Machine (SVM) in air quality prediction are significant. Firstly, SVM demonstrates better accuracy and reliability compared to linear prediction models, particularly in terms of Root Mean Square Error (RMSE), when predicting the Air Quality Index (AQI) and concentrations of pollutants such as PM2.5, PM10, NO, NO2, and NH3 [1] [2]. Additionally, SVM has the ability to handle complex data, effectively predicting PM2.5 concentrations in cities with intricate topography, which highlights its potential as a predictive model in other tropical areas [5]. Furthermore, SVM is utilized in air quality warning systems, where it has shown commendable performance in classifying air quality [4].

Support Vector Machine (SVM) has become a widely researched topic in air quality prediction and environmental data classification. As a powerful machine learning model, SVM is designed to handle non-linear and complex datasets by utilizing kernel functions to map data into higher-dimensional spaces, allowing for optimal separation between data classes. Research by Mogollón-Sotelo et al. [5] successfully applied the SVM model to predict ground-level ozone concentration, demonstrating the effectiveness of SVM in providing accurate prediction results by considering the complex atmospheric dynamics. Similarly, Balram et al. developed an SVM-based air quality warning system capable of classifying pollution levels in real-time, supporting quick decision-making for pollution mitigation.

In the context of particle pollution prediction, research by Widyarini and Purnomo [3] shows that the combination of SVM with advanced feature selection techniques can improve the accuracy of PM10 and PM2.5 predictions. This study highlights the importance of hyperparameter optimization in improving model performance. For comparison, compared SVM with other machine learning methods in a real-time air quality prediction system, finding that SVM provides a good balance between accuracy and computation time.

Additional research [6] [7]focuses on short-term prediction of PM10 and PM2.5 concentrations using Support Vector Regression, a variant of SVM specifically designed for regression. The results of this study indicate that SVM is capable of handling high temporal variations in air pollution data, making it an ideal choice for predictive implementation in large cities with fluctuating pollution dynamics. However, the existing challenges include the need to integrate SVM with other machine learning methods. [8][9] suggest an ensemble approach that combines SVM with artificial neural networks or neuro-fuzzy techniques to enhance prediction stability. [10] explored a dynamic SVM model that adjusts model parameters according to seasonal changes in air quality, extending the flexibility of the method for long-term applications.

Existing research shows the great potential of SVM in assisting the planning of air pollution control policies. With the continuous advancement of technology, the development of real-time integration and hybrid techniques will be the next important step in enhancing air quality prediction capabilities.

Development of Non-linear Models: With the increasing development of non-linear prediction models using machine learning, SVM can be used to quantify different dynamic processes related to air quality with shorter computation times compared to physical models [5]. Integration with Other Techniques: Further research can explore the integration of SVM with other soft computing techniques such as artificial neural networks and neuro-fuzzy systems to enhance prediction accuracy and stability [3].

SVM is a powerful and effective tool for predicting air quality, offering high accuracy and reliability in various contexts. With the continuous development of machine learning techniques, SVM can be integrated with other methods to improve air quality predictions and assist in decision-making related to air pollution control.

3. RESEARCH METHODS

In this research, the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology was employed as a structured approach to guide the process of analyzing and predicting air quality using Support Vector Machine (SVM). CRISP-DM is a widely accepted framework that provides a comprehensive process model for data mining projects and consists of six major phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Figure 1: Stages in CRISP-DM

Business Understanding

The first phase of CRISP-DM involves understanding the project objectives and requirements from a business perspective. In this research, the primary goal was to predict air quality levels, specifically the Air Quality Index (AQI) and concentrations of various pollutants, to inform public health policies and environmental regulations. This phase included identifying key stakeholders, such as environmental agencies and public health officials, and understanding their needs for accurate air quality predictions to enhance decision-making processes related to air quality management.

Data Understanding

Following the business understanding phase, the next step was to gather relevant data to support the analysis. This phase involved collecting historical air quality data, which encompassed various parameters such as PM2.5, PM10, nitrogen dioxide (NO2), carbon monoxide (CO), ozone (O3), and meteorological factors like temperature and humidity. The data sources included government environmental monitoring agencies, satellite data, and meteorological stations. During this phase, a preliminary analysis was conducted to assess the quality of the data, identify any missing values, and understand the distribution of different pollutants across various time frames and geographical locations.

Data Preparation

Data preparation is a crucial step in the CRISP-DM framework that involves cleaning and transforming the data into a suitable format for analysis. In this research, the collected data underwent several preprocessing steps, including handling missing values, normalizing the data, and encoding categorical variables. Additionally, exploratory data analysis (EDA) was conducted to visualize the relationships between different pollutants and to identify any trends or patterns that could inform the modeling process. This phase ensured that the data was ready for the modeling phase, enhancing the accuracy and efficiency of the predictions.

Modeling

The modeling phase involved selecting appropriate modeling techniques and building predictive models using SVM. Various configurations of SVM were tested, including different kernel functions (linear, polynomial, and radial basis function) and hyperparameter tuning to optimize model performance. The models were trained using a portion of the dataset while the rest was reserved for validation. Cross-validation techniques were employed to ensure that the models were robust and could generalize well to unseen data. This phase culminated in the selection of the best-performing SVM model based on evaluation metrics such as Root Mean Square Error (RMSE) and R-squared values.

Evaluation

Once the models were built, the evaluation phase assessed their performance in predicting air quality. The selected SVM model was evaluated against a separate test dataset to determine its accuracy and reliability. This phase involved analyzing the model's predictions against actual air quality measurements and calculating performance metrics. The evaluation not only helped in understanding the model's effectiveness but also provided insights into potential areas for improvement, such as feature selection or incorporating additional data sources.

Deployment

Finally, the deployment phase focused on integrating the predictive model into a user-friendly system for stakeholders. This involved creating a dashboard or application that displays real-time air quality predictions, alerts, and historical trends. The deployment phase also included the documentation of the model and its processes to ensure that stakeholders could effectively utilize the system for informed decision-making regarding air quality management.

The CRISP-DM framework provided a structured approach to conducting this research on air quality prediction using SVM. By following the phases of CRISP-DM, the research ensured a comprehensive analysis and developed a reliable predictive model that can support effective air quality management and contribute to public health initiatives. The methodology not only facilitated the technical aspects of the research but also aligned the project with the needs of stakeholders, making it a valuable tool for addressing air quality challenges.

4. RESULT AND DISCUSSION

The author implements numerous procedures when employing the CRISP-DM methodology, including:

Business Understanding

At the business understanding stage, researchers identify problems, determine business objectives, and understand the context related to air quality in Jakarta. The researchers analyze historical ISPU data, study regulations and policies related to air quality, and identify the stakeholders involved. The goal of this stage is to gain a comprehensive understanding of the issues and how analytical solutions can add value. The goal of this research is to construct a predictive ISPU model, utilizing the Support Vector Machine (SVM) algorithm, to aid in air quality decision-making in Jakarta, thereby fostering the growth of a more sustainable city. In other words, the researchers gather information and clearly define the problem before moving on to the next stage.

Data Understanding

This research utilizes data from the ISPU DKI Jakarta source, encompassing daily measurements of air pollutant concentrations, including PM10, PM2.5, SO2, CO, and O3, along with ISPU categories (GOOD, MODERATE, UNHEALTHY, etc.). At this stage, the researchers do exploration and analysis of the data to ascertain its characteristics, quality, and potential concerns. Researchers perform descriptive statistical analysis to provide a preliminary understanding of the data distribution for each pollutant variable. Researchers utilize visualizations, such histograms and box plots, to discern trends and outliers within the data. We performed additional data exploration by examining the correlation among pollutant variables through a correlation matrix and heatmap. The objective is to ascertain correlations among variables and discover characteristics that may serve as effective predictors for the model. The research results indicate a positive association between PM10 and PM2.5, suggesting that both pollutants tend to rise or fall concurrently. Researchers can ascertain suitable solutions for data preparation and modeling stages by a profound comprehension of the data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4626 entries, 0 to 4625
Data columns (total 11 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   tanggal   4626 non-null   object
 1   stasiun   4626 non-null   object
 2   pm10      4466 non-null   float64
 3   pm25      723 non-null    float64
 4   so2       4607 non-null   float64
 5   co        4618 non-null   float64
 6   o3        4621 non-null   float64
 7   no2       4618 non-null   float64
 8   max       4626 non-null   float64
 9   critical  4625 non-null   object
 10  categori  4626 non-null   object
dtypes: float64(7), object(4)
memory usage: 397.7+ KB
```

Figure 2: Dataset Info

Data Preparation

The data preparation phase plays a crucial role in the development of a machine learning model. At this stage, we carry out a series of procedures to cleanse, convert, and prepare the data for model training. The main focus at this stage is on tackling missing values and selecting relevant features. For instance, the team conducted descriptive statistical analysis of the data distribution of key pollutant variables, such as PM10, PM2.5, SO2, CO, and O3. We handled the dataset's missing values through imputation, substituting the mean value of each variable. The chosen imputation approach was based on the characteristics of the data and the premise that the missing values are either completely random (missing at random/MCAR) or random (missing at random/MAR).

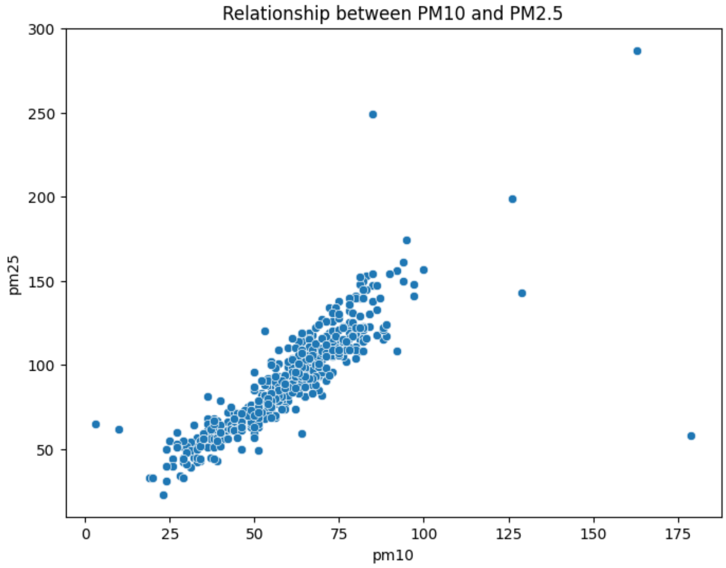

Figure 3: Relationship Pm10 and OM2.5

The team subsequently chose the features for the SVM model according to their importance and impact on the ISPU category. Visualizations like scatter plots (Figure 3) and correlation matrices revealed insights into the relationships among pollutants, especially the strong positive association between PM10 and PM2.5, which confirmed their role as key predictors in the model. Furthermore, specialized knowledge and previous evaluations of ISPU category distributions (Figure 4) contributed to identifying the most critical features: PM10, PM2.5, SO_2, CO, and O_3.

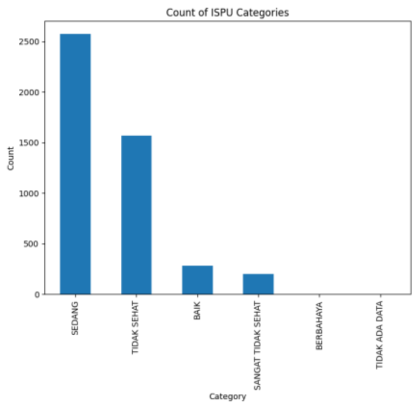

Figure 4 : Count ISPU

The analysis of time-series trends for pollutants such as PM10 (Figure 5) confirmed that the chosen features effectively represented both short-term variations and long-term trends. By conducting these steps to prepare the data, including addressing missing values and selecting the best features, the team ensured that the data used to build the SVM model was of high quality, consistent, and ready to produce an accurate and reliable model.

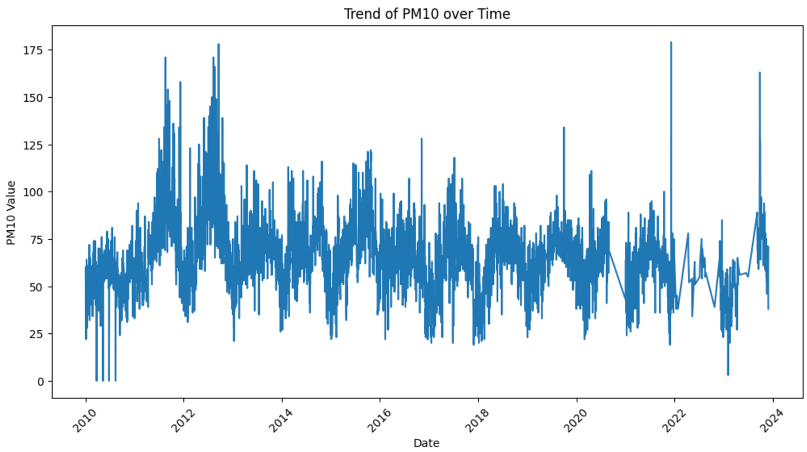

Figure 5: Trend PM 10

Modelling

During the Modelling phase, the Support Vector Machine (SVM) technique was selected for its capacity to manage non-linear data and generate resilient models. SVM works by finding the optimal hyperplane that separates the data into different classes. Prior to initiating the model training procedure, the prepared data is partitioned into training data (80%) and testing data (20%). The data partitioning is executed randomly via the train_test_split() function from the scikit-learn library to guarantee that both subsets accurately represent the entire dataset and to mitigate overfitting, a phenomenon where the model is excessively trained on the training data and fails to generalise effectively to new data.

A correlation study across features was performed utilising a heatmap to elucidate the linkages and strength of associations between predictor variables (PM10, PM2.5, SO2, CO, O3). This information elucidates how these features influence the prediction of ISPU categories and can serve as a foundation for feature selection or additional feature engineering.

Upon comprehending the association among features, the parameter tuning process is executed to identify the optimal model configuration. The Radial Basis Function (RBF) kernel was selected because of its efficacy with non-linear data. The C parameter (regularisation parameter) and gamma (kernel parameter) are optimised by grid search, which entails testing multiple combinations of parameter values and assessing the model's performance on the validation dataset. The objective of parameter tuning is to identify the optimal combination of parameter values that maximises model performance on the training data, as assessed by measures like accuracy.

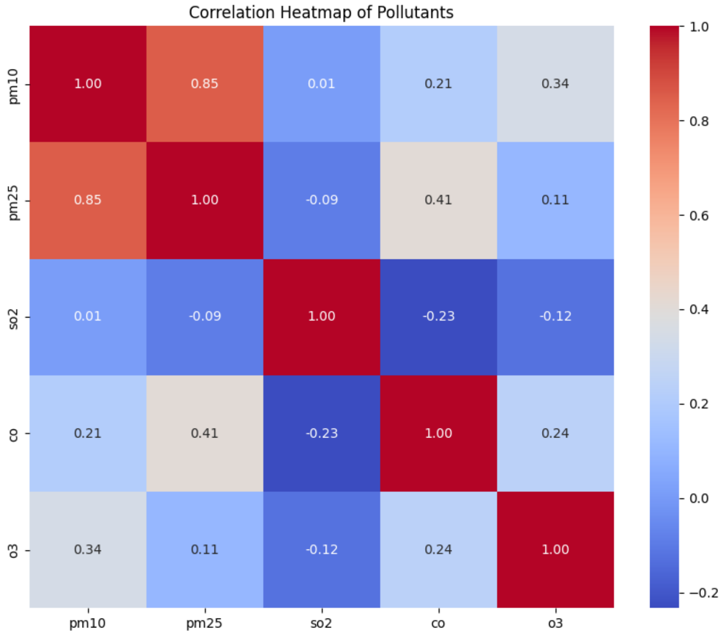

Figure 4: Correlation Heatmap of Pollutans

Evaluation

Multiple metrics assess the efficacy of the constructed SVM model during the evaluation phase, gauging its accuracy in predicting ISPU categories. The employed measures encompass accuracy, precision, recall, and F1-score. Accuracy quantifies the ratio of correct predictions relative to the entire dataset, whereas precision quantifies the ratio of true positive predictions relative to the total positive predictions. Recall quantifies the ratio of genuine positive predictions to all real positive instances, while the F1 score is the harmonic mean of precision and recall, offering a balance between these two metrics.

Moreover, the confusion matrix serves to illustrate the model's effectiveness in categorizing ISPU classifications. The confusion matrix shows the number of correct and incorrect predictions for each category, thus providing a more detailed picture of the types of errors made by the model.

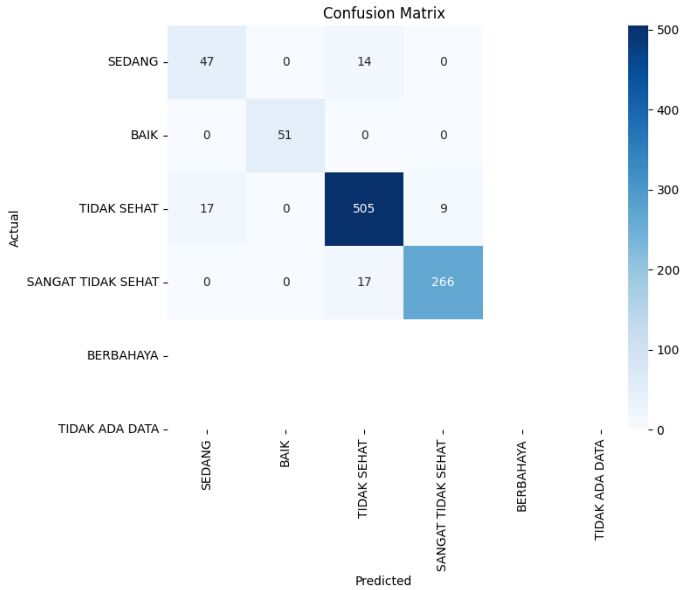

Figure 5: Confusion Matrix

The evaluation findings indicate that the SVM model, utilizing optimal parameters derived during the modeling phase, attained good accuracy in predicting ISPU categories. The model achieved an accuracy of 0.9384, accurately predicting 93.84% of the data, as illustrated in Figure 6. The model has excellent performance in categorizing the "SANGAT TIDAK SEHAT" category, achieving perfect scores (1.00) across precision, recall, and F1-score measures. The model accurately predicts the "SEDANG" and "TIDAK SEHAT" categories, demonstrating precision, recall, and F1-score values exceeding 0.94. The "BAIK" category exhibits marginally reduced precision, recall, and F1-score values, approximately 0.75, suggesting that the model may have challenges in accurately identifying this category.We selected the Support Vector Machine (SVM) technique during the modelling phase due to its ability to manage non-linear data and produce resilient models. Support Vector Machine (SVM) operates by identifying the best hyperplane that delineates distinct classes within the data. Before starting the model training procedure, we partition the prepared data into 80% training data and 20% testing data. The train_test_split() function from the scikit-learn library randomly executes the data partitioning to ensure both subsets accurately represent the entire dataset and to prevent overfitting, a phenomenon where the model overtrains on the training data and fails to generalize effectively to new data.

Accuracy: 0.9384449244060475

	precision	recall	f1-score	support
BAIK	0.73	0.77	0.75	61
SANGAT TIDAK SEHAT	1.00	1.00	1.00	51
SEDANG	0.94	0.95	0.95	531
TIDAK SEHAT	0.97	0.94	0.95	283
accuracy			0.94	926
macro avg	0.91	0.92	0.91	926
weighted avg	0.94	0.94	0.94	926

Figure 6: Model Evaluation

We performed a correlation study across features using a heatmap to elucidate the linkages and strength of associations between predictor variables (PM10, PM2.5, SO2, CO, O3). This information aids in comprehending how these features influence the prediction of ISPU categories and can serve as a foundation for feature selection or additional feature engineering.

After understanding the relationship between features, we execute the parameter optimisation process to determine the optimal model configuration. We selected the Radial Basis Function (RBF) kernel due to its effectiveness with non-linear data. Grid search optimises the C parameter (regularisation parameter) and gamma (kernel parameter) by testing multiple combinations of parameter values and evaluating the model's performance on the validation dataset. The objective of parameter tuning is to identify the optimal combination of parameter values that maximises model performance on the training data, assessed using measures like accuracy.

5. CONCLUSION

This research concludes that Support Vector Machine (SVM) is an effective method for predicting and classifying air quality. The results show that SVM is capable of handling complex and non-linear data, providing accurate and reliable prediction outcomes. SVM works better than linear prediction models when using historical datasets that include different pollutant parameters. It is especially good at being accurate and reliable. According to Figure 5, which shows the confusion matrix, SVM correctly sorts air quality categories 93.84% of the time. It does best with the "VERY UNHEALTHY" category, which has a precision, recall, and F1 score of 1.00. This indicates that the method could be applied in real-world scenarios, such as air quality warning systems. Additionally, Figure 6, which displays the model evaluation metrics, further confirms the model's robust performance, highlighting its strengths in categories such as 'TIDAK SEHAT' and 'SEDANG.'". Furthermore, this research highlights the challenges in developing non-linear models and emphasizes the importance of combining SVM with other techniques to enhance prediction accuracy. Overall, SVM has proven to be a powerful tool in air quality management, and with advancements in machine learning techniques, there is a tremendous opportunity to improve the effectiveness of this method in controlling air pollution in the future.

UCAPAN TERIMA KASIH

The author expresses profound appreciation to all parties who have provided support throughout this research. We are grateful for the valuable contributions of colleagues involved in the development and testing process, as well as for the constructive feedback that has helped refine the research findings. Every form of assistance, whether direct or indirect, is highly appreciated in supporting the achievement of this study's objectives.

REFERENCE

[1] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao, Y.-S. Wang, and C.-C. Kang, "Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting," Science of The Total Environment, vol. 651, pp. 230–240, Feb. 2019, doi: 10.1016/j.scitotenv.2018.09.111.

[2] M. Kulkarni, A. Raut, S. Chavan, N. Rajule, and S. Pawar, "Air Quality Monitoring and Prediction using SVM," in 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, IEEE, Aug. 2022, pp. 1–4. doi: 10.1109/ICCUBEA54992.2022.10010942.

[3] L. Widyarini and H. D. Purnomo, "Air Quality Prediction Using the Support Vector Machine Algorithm," Journal of Information Systems and Informatics, vol. 6, no. 2, pp. 652–661, Jun. 2024, doi: 10.51519/journalisi.v6i2.705.

[4] D. Balram, K.-Y. Lian, and N. Sebastian, "Air quality warning system based on a localized PM2.5 soft sensor using a novel approach of Bayesian regularized neural network via forward feature selection," Ecotoxicol Environ Saf, vol. 182, p. 109386, Oct. 2019, doi: 10.1016/j.ecoenv.2019.109386.

[5] C. Mogollón-Sotelo, A. Casallas, S. Vidal, N. Celis, C. Ferro, and L. Belalcazar, "A support vector machine model to forecast ground-level PM2.5 in a highly populated city with a complex terrain," Air Qual Atmos Health, vol. 14, no. 3, pp. 399–409, Mar. 2021, doi: 10.1007/s11869-020-00945-0.

[6] B.-Y. Kim, Y.-K. Lim, and J. W. Cha, "Short-term prediction of particulate matter (PM10 and PM2.5) in Seoul, South Korea using tree-based machine learning algorithms," Atmos Pollut Res, vol. 13, no. 10, p. 101547, Oct. 2022, doi: 10.1016/j.apr.2022.101547.

[7] Y.-C. Liang, Y. Maimury, A. H.-L. Chen, and J. R. C. Juarez, "Machine Learning-Based Prediction of Air Quality," Applied Sciences, vol. 10, no. 24, p. 9151, Dec. 2020, doi: 10.3390/app10249151.

[8] H. Zheng, Y. Cheng, and H. Li, "Investigation of model ensemble for fine-grained air quality prediction," China Communications, vol. 17, no. 7, pp. 207–223, Jul. 2020, doi: 10.23919/J.CC.2020.07.015.

[9] W.-I. Lai, Y.-Y. Chen, and J.-H. Sun, "Ensemble Machine Learning Model for Accurate Air Pollution Detection Using Commercial Gas Sensors," Sensors, vol. 22, no. 12, p. 4393, Jun. 2022, doi: 10.3390/s22124393.

[10] Q. Zhang, F. Fu, and R. Tian, "A deep learning and image-based model for air quality estimation," Science of The Total Environment, vol. 724, p. 138178, Jul. 2020, doi: 10.1016/j.scitotenv.2020.138178.