Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

Analysis of Air Pollution Standard Index Using Support Vector Machine Algorithm

Fitra Hidayat Lubis*, M Fakhriza, Raissa Amanda Putri

Faculty of Science and Technology, Computer Science Program, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1,*]fitra0701223081@uinsu.ac.id, [2]fakhriza@uinsu.ac.id, [3]raissa.ap@uinsu.ac.id
Correspondence Author Email: fitra0701223081@uinsu.ac.id
Submitted: **09/03/2026**; Accepted: **31/03/2026**; Published**: 31/03/2026**

Abstract−Air pollution is one of the major environmental problems in urban areas, including Medan City, Indonesia. The Air Pollution Standard Index (Indeks Standar Pencemar Udara / ISPU) data provided by the Environmental Agency is often difficult for the public to interpret due to its numerical format. This study aims to analyze and classify air quality using the Support Vector Machine (SVM) algorithm and present the results through data visualization. The dataset used in this research is secondary data obtained from the Environmental Agency of Medan City, including pollutant parameters such as PM10, PM2.5, SO_2, NO_2, CO, O_3, and HC. The research method follows a quantitative descriptive approach, including data preprocessing, ISPU calculation based on government regulations, classification using SVM, and visualization using graphical methods such as line charts, bar charts, and heatmaps. The results indicate that SVM is effective in classifying air quality categories into Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous. Additionally, visualization techniques improve the interpretability of air quality data, making it easier for stakeholders and the public to understand environmental conditions. This study contributes to decision support systems for environmental monitoring and public awareness.

Keywords: Air Pollution; ISPU; Support Vector Machine; Data Visualization; Machine Learning

1. INTRODUCTION

Air pollution has become a major environmental concern in urban areas around the world, including Medan City. The increasing number of motor vehicles, rapid industrialization, and continuous population growth contribute significantly to the rise of air pollutant emissions. Common pollutants such as particulate matter (PM10 and PM2.5), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3) are released into the atmosphere in large quantities [1]. These pollutants not only degrade environmental quality but also pose serious risks to human health, including respiratory illnesses, cardiovascular diseases, and reduced overall quality of life. In developing urban regions, the lack of effective monitoring and public awareness further exacerbates the negative impacts of air pollution, making it a critical issue that requires immediate attention [2].

The impact of air pollution extends beyond human health and directly affects environmental sustainability and economic stability. Elevated pollutant concentrations can damage vegetation, disrupt ecological balance, and contribute to climate-related issues such as global warming and acid rain [3]. In urban settings like Medan City, poor air quality also reduces visibility and accelerates infrastructure degradation. Furthermore, long-term exposure to polluted air can increase healthcare costs and decrease workforce productivity, thereby affecting economic development. These multidimensional impacts highlight the importance of implementing reliable monitoring systems and analytical tools to better understand and manage air quality conditions [4].

In Indonesia, air quality assessment is conducted using the Air Pollution Standard Index (ISPU), which is regulated by the Ministry of Environment and Forestry [5]. The ISPU system transforms complex pollutant concentration data into a standardized numerical index that represents air quality levels ranging from “Good” to “Hazardous.” This transformation is intended to simplify communication between environmental agencies and the public [6]. However, despite its standardization, ISPU data is still primarily presented in tabular or numerical form. This format often makes it difficult for non-expert users to interpret patterns, trends, and the severity of air pollution conditions over time, limiting its effectiveness as a public information tool [7].

To address these limitations, various computational approaches have been introduced to enhance the analysis of air quality data. Machine learning techniques, in particular, have gained significant attention due to their ability to process large datasets and identify complex patterns [4]. Among these methods, Support Vector Machine (SVM) is widely recognized for its effectiveness in classification tasks, especially when dealing with non-linear and high-dimensional data [8]. SVM works by constructing an optimal hyperplane that separates data into different categories with maximum margin, making it suitable for classifying air quality levels based on multiple pollutant parameters [9].

Several previous studies have demonstrated the effectiveness of SVM in air quality classification. Research conducted by Awaludin et al, 2025 reported that SVM achieved an accuracy of 93.47% in classifying air quality data, indicating its reliability in handling environmental datasets [10]. Similarly, Yudha et al, 2024 showed that SVM can process complex air pollution data with accuracy exceeding 90%, particularly when appropriate kernel functions are applied [11]. Pratama et al, 2021 further confirmed that SVM is capable of predicting air quality conditions with high reliability across multiple pollutant parameters. These findings suggest that SVM is a robust method for air quality analysis and prediction [3].

On the other hand, alternative approaches such as K-Means clustering have also been applied in air quality studies. Ridho and Mahalisa, 2023 utilized K-Means to group air quality data and achieved an accuracy of 77.78%

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

[12]. However, due to its unsupervised nature, K-Means lacks the ability to utilize labeled data effectively, resulting in lower classification performance compared to supervised methods like SVM. While these studies provide valuable insights into air quality analysis, most of them focus primarily on improving classification accuracy and algorithm performance [13]. However, there is still a significant research gap in integrating machine learning classification with comprehensive data visualization techniques that can enhance the interpretability and usability of air quality information for the general public and decision-makers [14].

In addition to classification limitations, the lack of intuitive visualization remains a major challenge in communicating air quality data [15]. Raw numerical outputs and statistical results are often difficult to interpret, especially for non-technical users. Without effective visualization, it becomes challenging to identify trends, detect anomalies, and understand temporal variations in air pollution levels [12]. Visual tools such as line charts, bar graphs, heatmaps, and dashboards can significantly improve data comprehension by presenting information in a more accessible and interactive manner. Therefore, combining machine learning techniques with visualization approaches is essential to bridge the gap between data analysis and practical application [16].

Based on these considerations, this study aims to apply the Support Vector Machine (SVM) algorithm for classifying ISPU data and to develop effective data visualization techniques for presenting air quality information. By integrating classification and visualization, this research seeks to provide a more comprehensive understanding of air pollution conditions in Medan City [17]. The expected outcome is a system that not only delivers accurate classification results but also enhances user understanding through intuitive visual representations. Ultimately, this study is expected to support decision-making processes, improve environmental monitoring, and increase public awareness regarding the importance of maintaining air quality [10].

2. RESEARCH METHODOLOGY

2.1 Research Type and Approach

This study uses a quantitative research type with a descriptive approach and the application of machine learning methods. The quantitative approach is applied because the data used consists of numerical values obtained from air pollutant measurements, which are processed using statistical and computational techniques. The descriptive approach is used to describe air quality conditions based on ISPU values and to present the results in a clear and structured manner. In addition, the machine learning approach using Support Vector Machine (SVM) is implemented to build a classification model capable of predicting air quality categories accurately. This combination allows the research to produce both analytical and predictive results [18].

2.2 Research Stages

This research was conducted systematically starting from data collection to the preparation of research results. The first stage is data collection, where air quality data is obtained from the Environmental Agency of Medan City. The collected data includes several pollutant parameters such as PM10, PM2.5, SO_2, CO, NO_2, O_3, and HC, which are used as the main variables in determining air quality conditions.

The next stage is data cleaning, which involves checking for missing values, incomplete data, and invalid records. This process is essential to ensure that the dataset used in the analysis is accurate, consistent, and suitable for further processing. After obtaining a clean dataset, the study proceeds to the ISPU calculation stage. At this stage, the Air Pollution Standard Index (ISPU) is calculated based on pollutant concentration values using the formula determined by the Minister of Environment and Forestry Regulation No. 14 of 2020.

Once the ISPU values are obtained, the next step is determining the ISPU category. The calculated index values are classified into several categories, namely Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous. This classification helps in simplifying the interpretation of air quality conditions and provides a clearer understanding of environmental status.

Furthermore, the data is presented through visualization techniques. The results of ISPU calculations and classifications are displayed in the form of tables and graphical representations to facilitate analysis and improve data readability. Visualization plays an important role in identifying patterns, trends, and comparisons among pollutant parameters over time.

The final stage is analysis and interpretation of the results. In this stage, the researcher analyzes ISPU trends, identifies dominant pollutant parameters, and evaluates the overall air quality condition in Medan City. The results of this analysis are then used to draw conclusions and provide insights that can support decision-making in environmental monitoring [19].

As part of the system design, a flowchart is also developed to illustrate the workflow of the system visually, from input to output. The flowchart aims to simplify the understanding of the process within the system, particularly in describing the working stages of the Support Vector Machine (SVM) algorithm used in this study. The flowchart of the system is presented in Figure 1.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

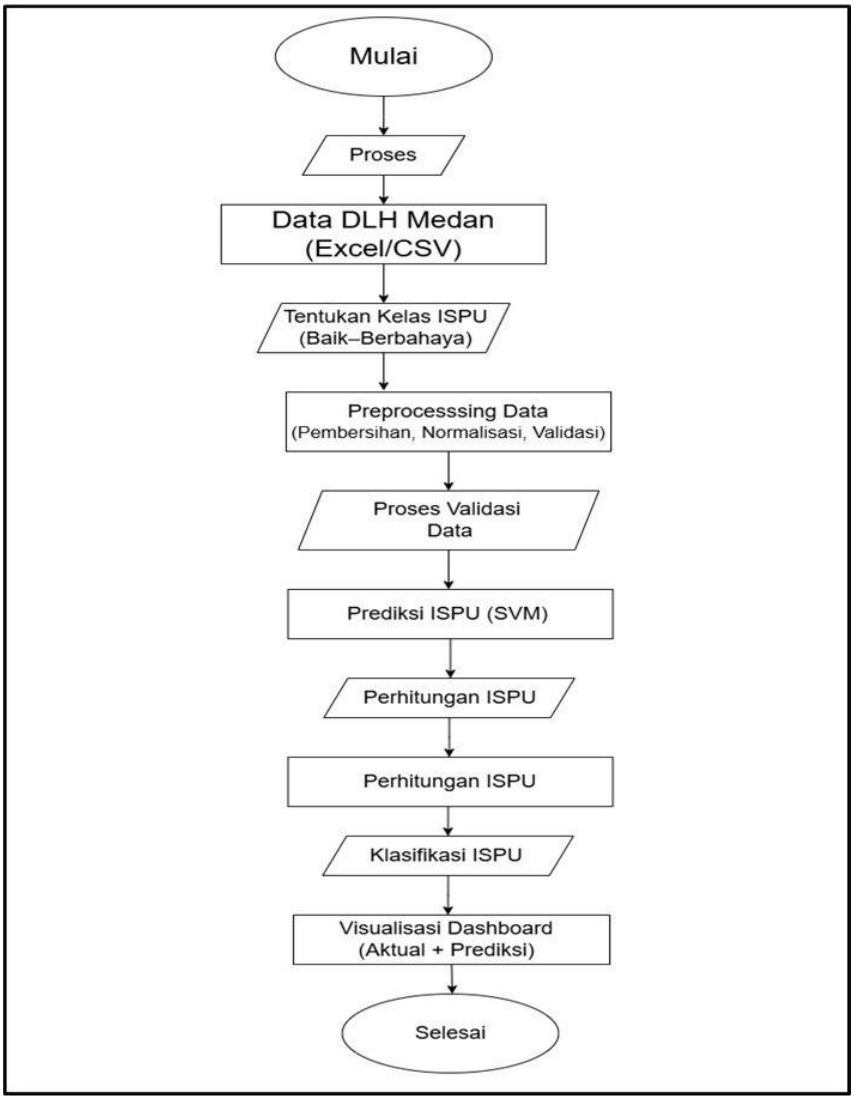

Figure 1. Flowchart system

2.3 Data Source

The dataset used in this study is secondary data obtained from the Environmental Agency of Medan City. The data represents the results of air quality monitoring from several stations in Medan City. The dataset includes hourly measurements of air pollutants such as Particulate Matter (PM10), Particulate Matter (PM2.5), Sulfur Dioxide (SO_2), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Ozone (O_3), and Hydrocarbon (HC). The data is considered valid and reliable because it is collected using standardized monitoring systems in accordance with government regulations. Therefore, the dataset reflects actual environmental conditions and can be used for accurate analysis and classification [20].

2.4 Data Processing

The data processing stage begins with importing air quality data used in this study. The dataset consists of 4,320 records collected from six monitoring locations in Medan City. The data spans from January 1, 2024, to January 30, 2024, with an hourly recording format (24 hours × 30 days × 6 locations). The parameters included in the dataset are PM10, PM2.5, SO_2, CO, NO_2, O_3, and HC, which are used as indicators in determining air quality levels.

Each record contains information such as location classification, date, hour, timestamp, and pollutant concentrations. The structured format of this dataset allows for detailed analysis of air pollution patterns across time and locations. The initial sample of the dataset is presented in Table 1, which shows the first five records of air quality data based on location, time, and PM10 values.

Table 1. Five Initial Air Quality Data

Location Classification	Date	Hour	Time	PM10
Medan Kota (City Center)	2024-01-01	0	00:00:00	62.944798
Medan Kota (City Center)	2024-01-01	1	01:00:00	67.492903

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

Location Classification	Date	Hour	Time	PM10
Medan Kota (City Center)	2024-01-01	2	02:00:00	25.621380
Medan Kota (City Center)	2024-01-01	3	03:00:00	50.806958
Medan Kota (City Center)	2024-01-01	4	04:00:00	44.509270

In addition to PM10 values, the dataset also includes other pollutant parameters such as PM2.5, SO_2, CO, NO_2, O_3, and HC. These parameters are essential in calculating the ISPU value and analyzing air quality conditions. A sample of these pollutant values is shown in Table 2, which presents the corresponding parameter values for the same observation period.

Table 2. ISPU Parameter Data Range

No	PM2.5	SO_2	CO	NO_2	O_3	HC
0	28.238580	40.128700	10.558035	35.833112	55.849874	57.918465
1	25.456415	39.069005	5.552187	33.887870	61.848725	22.714135
2	24.676785	23.390662	7.511903	30.172598	41.004973	56.773740
3	29.967237	19.238538	5.348156	38.731750	44.297481	45.787036
4	26.949773	27.534797	11.387741	37.686623	45.472842	50.291253

The dataset covers a time range from 2024-01-01 00:00:00 to 2024-01-30 23:00:00, ensuring continuous hourly observations. This data processing stage is crucial because it provides a structured and reliable dataset that will be used in further stages, including data cleaning, ISPU calculation, classification using SVM, and data visualization.

2.5 ISPU Calculation

The ISPU value is calculated based on the formula established by the Ministry of Environment and Forestry. This formula converts pollutant concentration values into a standardized index that represents air quality levels. The calculation process uses variables such as measured concentration, upper and lower limits of concentration, and ISPU threshold values. The resulting ISPU values are then used as the basis for determining air quality categories. This step is essential in transforming raw environmental data into meaningful information that can be used for further analysis.

2.6 Data Visualization

Data visualization is used to present the results of analysis in a more informative and understandable way. Various visualization techniques are applied, including line charts to display trends over time, bar charts to compare pollutant values, heatmaps to identify patterns, and pie charts to show the proportion of air quality categories.These visualization methods help transform complex numerical data into visual insights that are easier to interpret, especially for decision-makers and the general public [21].

3. RESULT AND DISCUSSION

3.1 Data Processing Stages

The data processing in this study was conducted through several systematic stages, including data input, data cleaning, and ISPU calculation. Initially, the raw dataset obtained from the Environmental Agency of Medan City was imported into the processing system using Google Colaboratory. The Python programming language was utilized with supporting libraries such as Pandas for data manipulation and Matplotlib for visualization. This environment enables efficient handling of structured datasets and supports reproducible analysis.

During the data cleaning stage, the dataset was examined to identify missing values, invalid data, and outliers. The results showed that a small portion of the data contained anomalies, particularly in parameters such as PM10 and PM2.5. Missing values were handled using interpolation techniques, while extreme values were retained if they represented actual environmental phenomena. This approach ensures that the dataset maintains its real-world characteristics while minimizing noise. The comparison between labeled and unlabeled data is presented in Figure 2.

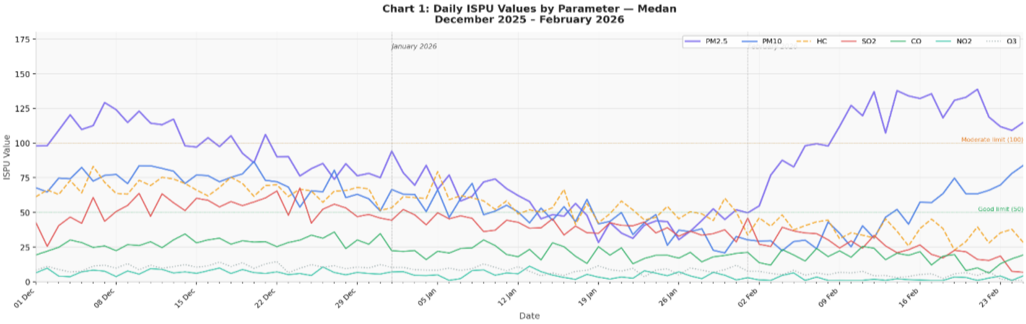

Figure 2. Displaying a multi-line graph of data processing results

ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

Figure 2, illustrates the comparison between labeled and unlabeled datasets in both time series and distribution perspectives. The time series plot shows that unlabeled data tends to exhibit higher volatility, with sharp fluctuations across observations. In contrast, labeled data appears smoother and more consistent, indicating that preprocessing and labeling reduce noise and improve data stability. The distribution plot further confirms this observation, where unlabeled data is widely dispersed, while labeled data is more concentrated around specific value ranges. This suggests that labeled data has better structure and is more suitable for machine learning processes. Overall, this visualization highlights the importance of preprocessing in enhancing data quality.

Following data cleaning, the ISPU calculation was performed based on the formula defined in the Regulation of the Minister of Environment. The ISPU value is calculated for each pollutant parameter, and the highest sub-index value is selected as the final ISPU. The results show that the average ISPU falls within the moderate category, indicating a relatively acceptable but slightly polluted air quality condition.

3.2 Determination of Air Quality Categories Using SVM

The classification of air quality was carried out using the Support Vector Machine (SVM) algorithm with the Radial Basis Function (RBF) kernel. The input features consist of pollutant parameters such as PM10, PM2.5, SO_2, CO, NO_2, O_3, and HC, while the output corresponds to the ISPU categories: Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous.

The dataset was divided into training and testing data with a ratio of 80:20. The model training process aimed to find the optimal hyperplane that separates data into different classes with maximum margin. The RBF kernel was chosen because it can transform non-linear data into a higher-dimensional space, allowing better separation between classes.

The evaluation results indicate that the model achieved perfect performance. All evaluation metrics, including accuracy, precision, recall, and F1-score, reached 100%. This means that the model successfully classified all testing data without any misclassification.

This high performance is influenced by the quality of the dataset and the clear separation between classes. Additionally, the RBF kernel plays a crucial role in capturing complex relationships between pollutant parameters, enabling the model to produce highly accurate classification results.

3.3 Data Description

The dataset used in this study was obtained from the Environmental Agency of Medan City through the SPKUA/AQMS (Air Quality Monitoring System), which continuously records air pollutant concentrations from several monitoring stations. The data covers the period from December 2025 to January 2026, a timeframe intentionally selected to capture short-term variations in air quality, particularly during the transition between the end and beginning of the year. This period is important because it may reflect changes in human activities, traffic density, and meteorological conditions that influence air pollution levels.

The dataset consists of 9,961 observations in the form of hourly time series data, meaning that measurements were recorded every hour over the study period. This high temporal resolution allows for detailed analysis of daily fluctuations, peak pollution hours, and temporal patterns such as diurnal cycles. For example, pollutant concentrations may increase during morning and evening rush hours due to traffic emissions, while lower concentrations may occur during nighttime when human activity decreases.

In terms of structure, the dataset includes both categorical and numerical variables. The categorical variables consist of time-related attributes such as hour (0–23), day, and month, as well as the type of pollutant parameter being measured. These variables are useful for grouping and analyzing patterns across different time intervals. Meanwhile, the numerical variable represents the concentration values of pollutants, measured in units such as $\mu g/m^3$ for particulate matter and ppm for gaseous pollutants. These numerical values are essential for calculating the ISPU index and performing statistical analysis.

The air quality parameters analyzed in this study include PM10, PM2.5, SO_2, CO, NO_2, O_3, and HC, which are commonly used indicators in air quality assessment. PM10 and PM2.5 represent particulate matter with diameters less than 10 µm and 2.5 µm, respectively, and are particularly harmful to human health because they can penetrate the respiratory system. Gaseous pollutants such as CO and NO_2 are primarily associated with vehicle emissions, while SO_2 is commonly linked to industrial activities. O_3 (ozone) is a secondary pollutant formed through photochemical reactions in the atmosphere, and HC (hydrocarbons) represents volatile organic compounds that contribute to air pollution formation.

Based on the data inspection process, the dataset was found to be complete and consistent, with no missing values or incomplete records. This indicates a high level of data reliability and eliminates the need for additional preprocessing techniques such as imputation. The completeness of the dataset ensures that subsequent analysis, including ISPU calculation, classification using SVM, and visualization, can be performed accurately and without bias. Overall, the quality and structure of the dataset make it highly suitable for modeling and analyzing air quality conditions in Medan City.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

3.4 Data Visualization Results

3.4.1 ISPU Trend (Line Chart)

The visualization of ISPU trends is presented in Figure 3.

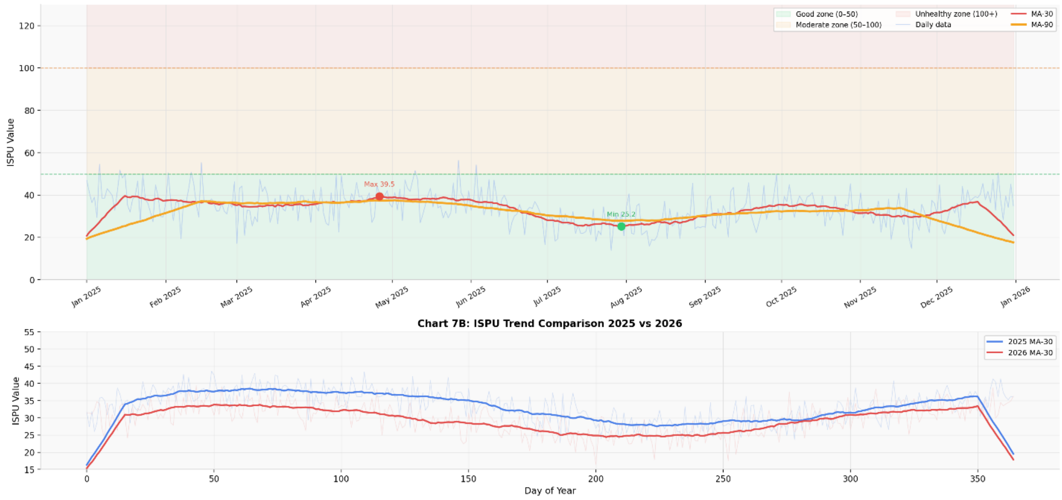

Figure 3. ISPU Trend Visualization

Figure 3 illustrates the temporal trend of ISPU values over the observation period using a line chart combined with moving average analysis. The thin line represents daily ISPU values, which show short-term fluctuations caused by dynamic environmental and anthropogenic factors. Meanwhile, the thicker lines represent moving averages (e.g., 30-day and 90-day), which smooth out short-term variations and highlight long-term trends and seasonal patterns.

From the visualization, it can be observed that ISPU values fluctuate significantly on a daily basis, indicating variability in pollutant concentrations influenced by traffic intensity, industrial emissions, and weather conditions. However, despite these fluctuations, there is a noticeable increasing trend in ISPU values over time, suggesting a gradual decline in air quality.

In addition, seasonal patterns are clearly visible. ISPU values tend to rise during certain periods, particularly in dry seasons, where lower rainfall reduces pollutant dispersion and increases pollutant accumulation in the atmosphere. Conversely, during rainy periods, pollutant concentrations tend to decrease due to the washing effect of precipitation. The comparison between 2025 and 2026 further confirms that the average ISPU has increased, indicating worsening air quality conditions over time.

3.4.2 Category Distribution (Bar Chart)

The distribution of ISPU categories is shown in Figure 4.

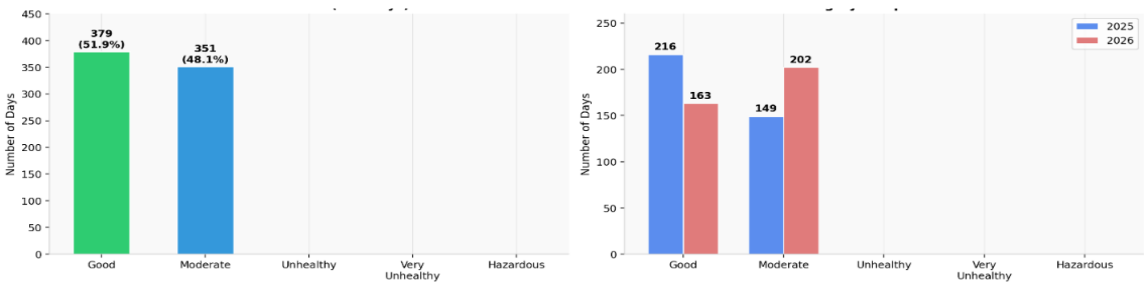

Figure 4. ISPU Category Distribution

Figure 4 presents the distribution of air quality categories in the form of a bar chart, showing the total number of days classified into each ISPU category. The categories include Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous. The results indicate that the "Good" category dominates, followed by the "Moderate" category, while higher pollution categories are either minimal or not present during the study period. This suggests that overall air quality is still within acceptable limits but shows signs of gradual degradation. To provide a clearer comparison between years, the distribution of ISPU categories is summarized in Table 3.

Table 3. Annual Comparison of ISPU Categories

Category	2025	2026	Change
Good	216	163	-53
Moderate	149	202	+53

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

Category	2025	2026	Change
Unhealthy	0	0	0
Very Unhealthy	0	0	0
Hazardous	0	0	0

Based on Table 3, there is a significant decrease in the number of “Good” days by 53 days, accompanied by an increase of 53 days in the “Moderate” category. This inverse relationship clearly indicates a shift in air quality conditions from Good to Moderate, suggesting a gradual decline in environmental quality. Although no days fall into the “Unhealthy,” “Very Unhealthy,” or “Hazardous” categories, the increase in Moderate conditions should be considered an early warning sign. This trend may be associated with factors such as increased urbanization, higher traffic emissions, and industrial activities. Therefore, the bar chart visualization, supported by the quantitative comparison in Table 4.1, provides a comprehensive understanding of how air quality conditions evolve over time. This information is particularly valuable for policymakers, as it highlights the need for preventive measures before more severe pollution levels occur.

3.4.3 Category Proportion (Pie Chart)

The proportion of ISPU categories is illustrated in Figure 5.

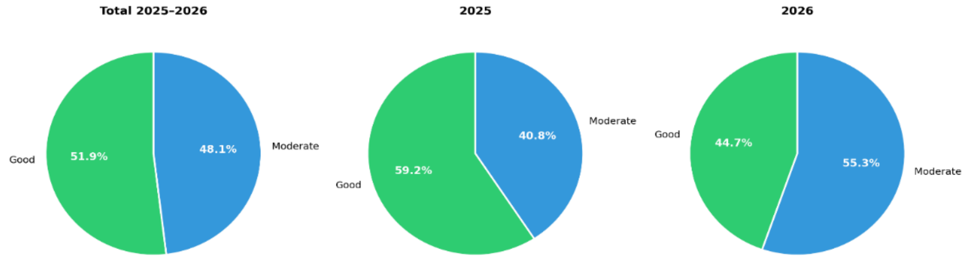

Figure 5. ISPU Category Proportion

Figure 5, presents the percentage distribution of air quality categories using pie charts. The visualization is divided into three parts: overall proportion, proportion for 2025, and proportion for 2026. From the overall distribution, it can be seen that the “Good” category still occupies the largest portion. However, when analyzed per year, a clear shift occurs. In 2025, the “Good” category is dominant, whereas in 2026, the “Moderate” category becomes more prominent. This shift indicates a gradual decline in air quality, where conditions that were previously classified as “Good” are now more frequently categorized as “Moderate.” Although the air quality is not yet at a dangerous level, this trend suggests increasing pollution levels that may pose risks in the future if not properly managed. The pie chart effectively highlights relative proportions, making it easier to identify dominant categories and observe changes in air quality composition over time.

3.4.4 Parameter Pattern (Heatmap)

The heatmap visualization is presented in Figure 6.

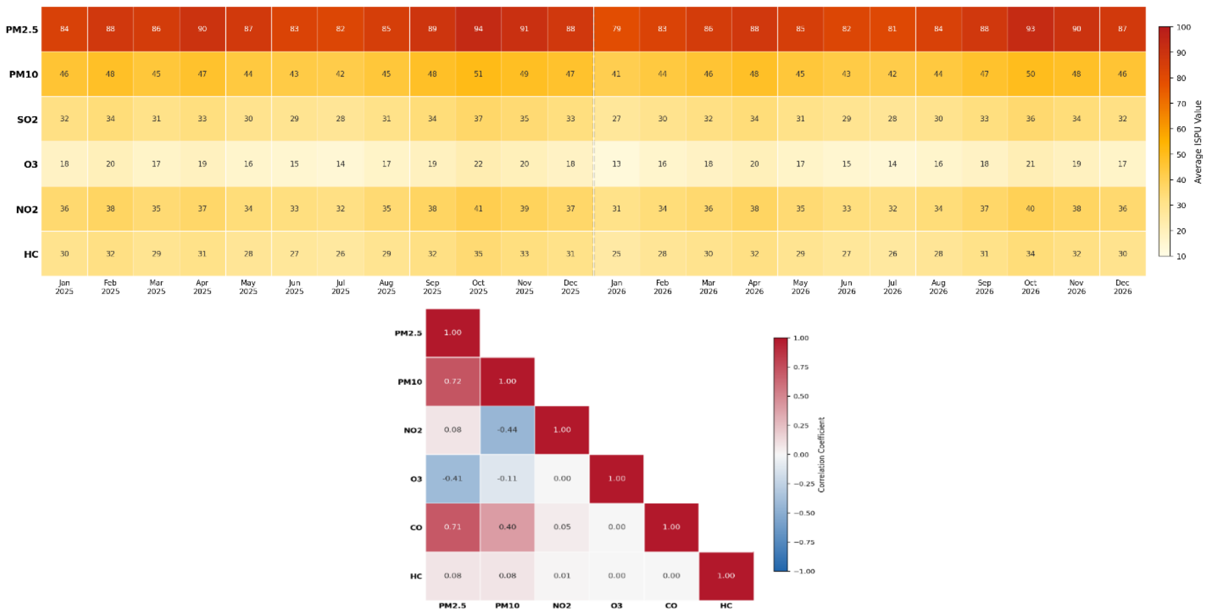

Figure 6. Heatmap of Parameter Concentration and Correlation

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

Figure 6, consists of two heatmaps that provide deeper insights into pollutant behavior and relationships. The first heatmap shows the monthly average concentration of pollutant parameters, where color intensity represents concentration levels (lighter colors indicate lower values, while darker colors indicate higher values). From this visualization, it can be observed that certain pollutants, particularly PM10 and PM2.5, exhibit higher concentrations during specific months. This indicates the presence of seasonal patterns, which may be influenced by meteorological factors such as temperature, humidity, wind speed, and rainfall. The second heatmap illustrates the correlation between pollutant parameters. The correlation values range from -1 to +1, where positive values indicate that two parameters increase together, while negative values indicate inverse relationships. Strong positive correlations are observed between:

a. PM10 and PM2.5
b. PM10 and O_3
c. PM2.5 and CO

These relationships suggest that these pollutants may originate from similar sources, such as vehicle emissions, industrial activities, and combustion processes. The presence of strong correlations also indicates that certain pollutants can act as indicators of overall air quality conditions. This heatmap is particularly useful for identifying dominant pollutants and understanding the interactions between different air quality parameters, which is essential for environmental monitoring and policy development.

3.5 Discussion

The results of this study reveal several important findings regarding air quality conditions in Medan City and the effectiveness of the Support Vector Machine (SVM) model in classifying ISPU data. Overall, the visualization and classification results consistently indicate a gradual decline in air quality, as evidenced by the increasing ISPU trend and the shift in dominant categories from "Good" to "Moderate." Although the values are still within acceptable limits, this shift is an early warning signal of increasing pollution levels that should not be overlooked. From the temporal analysis, the ISPU trend shows a fluctuating but increasing pattern over time. This suggests that air pollution in Medan is influenced not only by random daily variations but also by systematic factors such as urban growth, increased vehicle emissions, and industrial activities. The presence of seasonal patterns further strengthens this interpretation, where higher ISPU values tend to occur during dry periods. This aligns with environmental theory, where reduced rainfall limits pollutant dispersion, leading to higher pollutant accumulation in the atmosphere.

The distribution and proportion analysis provide additional insight into the structural shift in air quality conditions. The decrease in "Good" category days and the corresponding increase in "Moderate" days indicate that air quality is not deteriorating abruptly but rather gradually transitioning to a lower quality state. This pattern is critical from a policy perspective, as gradual degradation is often less noticeable but can have long-term cumulative impacts on public health. The heatmap analysis highlights that PM10 and PM2.5 are the dominant pollutants, showing both high variability and strong correlations with other parameters. This finding is consistent with previous studies, where particulate matter is identified as a primary contributor to air pollution in urban areas. The strong correlation between PM10 and PM2.5 suggests a common emission source, most likely vehicular emissions, road dust, and industrial combustion processes. Additionally, the correlation between particulate matter and gases such as CO and O_3 indicates complex atmospheric interactions, including photochemical reactions.

From a modeling perspective, the SVM algorithm demonstrates exceptional performance, achieving 100% accuracy, precision, recall, and F1-score. While this indicates that the model is highly effective in classifying the dataset, it is important to interpret these results critically. Perfect accuracy may suggest that the dataset has well-separated classes or relatively low complexity, but it may also indicate the possibility of overfitting, especially if the dataset is limited or not sufficiently diverse. Therefore, further validation using larger datasets or cross-validation techniques is recommended to ensure the model's generalizability. Compared to previous studies, the results of this research confirm that SVM is a robust method for air quality classification, particularly when combined with non-linear kernels such as RBF. However, the main contribution of this study lies not only in classification accuracy but also in the integration of data visualization techniques, which enhance interpretability. Visualization methods such as line charts, bar charts, pie charts, and heatmaps provide a more intuitive understanding of air quality patterns, making the results more accessible for decision-makers and the general public.

In terms of practical implications, the findings of this study highlight the need for preventive environmental policies. Although air quality is עדיין within safe limits, the observed declining trend suggests that proactive measures are necessary to prevent further deterioration. These measures may include stricter emission regulations, promotion of public transportation, and improved urban planning to reduce pollution sources. However, this study has several limitations. First, the dataset covers a relatively short time period, which may not fully capture long-term trends and extreme pollution events. Second, the analysis focuses primarily on pollutant concentration data without incorporating meteorological variables such as temperature, humidity, and wind speed, which play a significant role in air pollution dynamics. Future research is recommended to include these variables and explore more advanced models such as deep learning or hybrid approaches for improved prediction accuracy.In conclusion, this study demonstrates that the combination of SVM classification and data visualization provides a powerful approach for analyzing air quality. The results not only offer high classification accuracy but also deliver meaningful insights into pollution patterns, trends, and potential risks, thereby supporting more informed environmental decision-making.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

4. CONCLUSION

Based on the results of this study, it can be concluded that the analysis of the Air Pollution Standard Index (ISPU) in Medan City using the Support Vector Machine (SVM) method, combined with data processing and visualization techniques, is able to provide a clearer and more comprehensive understanding of air quality conditions. The dataset obtained from the Environmental Agency (DLH) of Medan City, which includes key pollutant parameters such as PM10, PM2.5, SO_2, CO, NO_2, O_3, and HC in accordance with the Regulation of the Minister of Environment and Forestry No. 14 of 2020, shows that the air quality during the study period is predominantly in the “Good” (51.9%) and “Moderate” (48.1%) categories, indicating relatively safe conditions despite a noticeable shift toward moderate levels in the following period. The implementation of SVM demonstrates excellent performance with accuracy, precision, recall, and F1-score reaching 100%, confirming its effectiveness in classifying air quality categories based on pollutant data, while visualization techniques such as line charts, bar charts, pie charts, and heatmaps successfully present the data in a more informative and easily interpretable manner. Therefore, it is recommended that future studies utilize longer observation periods and more monitoring locations to improve representativeness, explore alternative machine learning methods such as Random Forest, Artificial Neural Networks, or deep learning for performance comparison, integrate meteorological variables to enhance predictive capability, and develop real-time web or mobile-based visualization dashboards so that air quality information can be accessed more widely by the public and policymakers, ultimately supporting more effective environmental monitoring and air pollution control strategies in urban areas.

REFERENCES

[1] S. Bhattacharya and S. Shahnawaz, "Using Machine Learning to Predict Air Quality Index in New Delhi," ArXiv, vol. abs/2112.05753, 2021, doi: 10.48550/arXiv.2112.05753.
[2] C. N. Wijaya, H. Gunawan, and S. Wahyudi, "Pembuatan Media Promosi Katalog Produk Elektronik Menggunakan Teknologi Augmented Reality," Jurnal Protekinfo, vol. 8, no. 2, pp. 2–6, 2021, doi: 10.30656/protekinfo.v8i2.5028.
[3] A. Pratama, N. Hariyanto, A. Maulana, D. Primanda, A. Hidayati, and W. Prayitno, "Prediction of Air Pollution Standard Index (ISPU) Categories in DKI Jakarta Using the Gradient Boosting Algorithm," Advances in Intelligent Computing and Pattern Recognition (AICP), vol. 1, no. 1, pp. 402–416, 2021, doi: 10.52802/aicp.v1i1.1354.
[4] A. N. Fachmi, A. P. Sari, F. R. Ramadhan, R. D. Patria, S. Pudjiati, and F. N. Hassan, "Analisis Klasifikasi Indeks Standar Pencemaran Udara Jakarta Tahun 2025 Menggunakan Algoritma Random Forest," Jurnal Informatika dan Teknik Elektro Terapan (JITET), vol. 14, no. 1, pp. 529–537, 2025, doi: 10.23960/jitet.v14i1.8477.
[5] D. Salsabila and N. F. Utami, "Klasifikasi dan Pemetaan Spasial Kualitas Udara Berbasis ISPU Menggunakan Support Vector Machine," Journal of Science and Technology Letters (JSTL), vol. 17, no. 2, pp. 167–187, 2025, doi: 10.20885/jstl.vol17.iss2.art6.
[6] E. Fitri and A. Saryoko, "Determining Air Quality Influential Parameters Using Machine Learning Techniques," Journal of Dinda: Data Science, Information Technology, and Data Analytics, vol. 4, no. 2, pp. 105–114, 2024, doi: 10.20895/dinda.v4i2.1567.
[7] R. P. Roris, A. Saputra, A. Fahrizal, S. Susilowati, H. Rianto, and Y. Nuryamin, "Prediksi ISPU Jakarta Menggunakan Random Forest," Journal Automation Computer Information System (JACIS), vol. 5, no. 2, pp. 293–305, 2025, doi: 10.47134/jacis.v5i2.139.
[8] A. Gangwar, S. Singh, R. Mishra, and S. Prakash, "The State-of-the-Art in Air Pollution Monitoring and Forecasting Systems Using IoT, Big Data, and Machine Learning," ArXiv, vol. abs/2304.09574, 2025, doi: 10.48550/arXiv.2304.09574.
[9] S. Arti and E. Suherlan, "Evaluasi Kinerja Machine Learning dalam Memprediksi Kemampuan Adaptasi Mahasiswa pada Lingkungan Pembelajaran Daring," Jurnal Pustaka AI, vol. 5, no. 1, pp. 50–57, 2025, doi: 10.55382/jurnalpustakaai.v5i1.901.
[10] M. Awaludin, F. Risyda, and A. G. Gani, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Kualitas Udara di Wilayah Halim Perdanakusuma," Jurnal Mahasiswa Informatika dan Desain, vol. 2, 2025, doi: 10.35968/pgjw6t89.
[11] Y. A. Sani and B. Wasito, "Analisis Komparasi Algoritma Support Vector Machine dan K-Nearest Neighbor pada Klasifikasi Kualitas Udara Kota Jakarta," Grebuci: Jurnal Komputasi dan Rekayasa, vol. 2, no. 1, pp. 55–72, 2024, doi: 10.46806/grebuci.v2i1.1757.
[12] I. I. Ridho and G. Mahalisa, "Analisis Klasifikasi Dataset Indeks Standar Pencemaran Udara (ISPU) di Masa Pandemi Menggunakan Algoritma Support Vector Machine (SVM)," Technologia: Jurnal Ilmiah, vol. 14, no. 1, pp. 38–41, 2023, doi: 10.31602/tji.v14i1.8005.
[13] R. C. Irjayana, A. Fadlil, and R. Umar, "Air Quality Index Classification: Feature Selection for Improved Accuracy with Multinomial Logistic Regression," Jurnal Teknik Informatika (JUTIF), vol. 6, no. 5, pp. 3265–3279, 2025, doi: 10.52436/1.jutif.2025.6.5.5155.
[14] A. Patil and P. Patil, "Air Quality Index Prediction Using Machine Learning," International Journal of Advanced Computer Technology and Engineering (IJACTE), vol. 14, no. 1, 2025, doi: 10.65521/ijacte.v14i1.211.
[15] S. Arisa, "Penerapan Metode Support Vector Regression (SVR) dalam Memprediksi Indeks Standar Pencemar Udara (ISPU) di Kota Makassar," Variansi: Journal of Statistics and Its Application, vol. 7, no. 2, 2025, doi: 10.35580/variansiunm400.
[16] V. Wijaya, "Klasifikasi Kualitas Index Udara Dunia dengan Metode Artificial Neural Network dan Support Vector Machine Kernel RBF," Jurnal Protekinfo, vol. 19, no. 2, pp. 103–110, 2024, doi: 10.30656/protekinfo.v8i2.5028.
[17] I. G. A. N. Lestari and I. K. A. Aryanto, "Peningkatan Akurasi Klasifikasi Kualitas Udara melalui Oversampling dengan Metode Support Vector Machine dan Random Forest," Jurnal Sistem dan Informatika (JSI), vol. 18, no. 1, pp. 1–9, 2023, doi: 10.30864/jsi.v18i1.596.
[18] R. Rabiei, P. Bastani, H. Ahmadi, S. Dehghan, and S. Almasi, "Developing Public Health Surveillance Dashboards: A Scoping Review on the Design Principles," BMC Public Health, pp. 1–15, 2024, doi: 10.1186/s12889-024-17841-2.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2761−2770
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9506

[19] R. Firdaus, H. Habibie, and Y. Rizki, "Implementasi Algoritma Random Forest untuk Klasifikasi Pencemaran Udara di Wilayah Jakarta Berdasarkan Jakarta Open Data," Jurnal Fasilkom, vol. 14, no. 2, pp. 520–525, 2021, doi: 10.37859/jf.v14i2.7669.

[20] J. Han, W. Zhang, H. Liu, and H. Xiong, "Machine Learning for Urban Air Quality Analytics: A Survey," ArXiv, vol. abs/2310.09620, 2021, doi: 10.48550/arXiv.2310.09620.

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