Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

Application of Support Vector Machine with Kriging Interpolation for Rainfall Prediction in Java Island

Brian Dimas Purwanto*, Sri Suryani Prasetiyowati, Yuliant Sibaroni

School of Informatics, Informatics, Telkom University, Bandung, Indonesia
Email: briandimaspurwanto@student.telkomuniversity.ac.id, srisuryani@telkomuniversity.ac.id, yuliant@telkomuniversity.ac.id
Correspondence Author Email: briandimaspurwanto@student.telkomuniversity.ac.id
Submitted: **06/08/2024**; Accepted: **10/09/2024**; Published**: 12/04/2024**

Abstract−Rainfall is one of the crucial meteorological elements that can significantly impact human life. Accurate rainfall prediction is essential for effective natural resource planning and management across various regions, especially in Java Island, which is one of the most densely populated areas in Indonesia. However, predicting rainfall distribution in this region is challenging due to its complex climate patterns. This study aims to address this challenge by developing a rainfall distribution prediction model for Java Island using Support Vector Machine (SVM). The model incorporates time-based feature expansion to enhance the predictive capability of the SVM. Additionally, the method is combined with Kriging interpolation to accurately classify the spatial distribution of rainfall across Java Island. The model was trained and tested using monthly data from 27 meteorological stations, covering the period from January 1, 2020, to March 31, 2022. The results demonstrate that the model's performance exceeds 90%, indicating its effectiveness in predicting future rainfall distribution classifications. The contribution of this research lies in providing insights into feature expansion techniques in machine learning, refining predictive models applied in meteorology and environmental management, and addressing the complexity of rainfall prediction in densely populated regions.

Keywords: Support Vector Machine; time-based feature expansions; rainfall; classification prediction; interpolation krigging;

1. INTRODUCTION

One of the most important meteorological elements with significant impacts on human life is rainfall. Adequate rainfall is crucial for supporting agricultural activities and managing water resources. However, excessive rainfall can lead to natural disasters such as floods. Accurate rainfall prediction is essential for effective natural resource planning and management across various regions, particularly in Java Island, which is one of the most densely populated areas in Indonesia. Rainfall prediction in tropical regions like Java Island faces several challenges due to high climate variability and complex weather patterns [1], [2], [3]. Thus, there is a critical need for precise rainfall classification and prediction for each region in Java Island to better manage resources and mitigate disaster risks.

To address these challenges, various statistical and machine learning methods have been developed, including the use of Support Vector Machines (SVM). SVM has emerged as a promising approach for predicting rainfall due to its ability to handle nonlinear and multidimensional data. Recent studies highlight the effectiveness of SVM and its hybrid models in meteorological predictions, particularly for rainfall classification [4], [5], [6]. Liu et al. discuss enhancing rainfall prediction accuracy by integrating SVM with Kriging, illustrating the benefits of combining these methods [7]. Research [8], [9], [10], [11] has employed hybrid SVM models to enhance rainfall prediction accuracy in tropical climates. Spatial variability in rainfall data requires effective interpolation techniques. Kriging is a highly effective interpolation method for spatial data analysis. Wang, Xu, and Zhao provided a comparative analysis of Kriging regression and SVM, discussing the advantages of using Kriging in spatial data interpolation [12]. Kumar et al. demonstrated that accuracy improves when Kriging interpolation is used in conjunction with machine learning models [13], [14], [15].

This research aims to develop a more accurate time-based rainfall distribution classification prediction model for Java Island by integrating SVM classification methods with time-based feature expansion and Kriging interpolation. The primary purpose of this research is to improve the accuracy of rainfall prediction models by addressing the spatial and temporal complexities specific to Java Island. While previous studies have combined SVM with various interpolation techniques, our approach uniquely focuses on expanding time-based features, which has not been extensively explored in prior research. By incorporating time-based feature expansion, we enhance the model's ability to capture temporal patterns in rainfall data, thus improving the prediction accuracy for future rainfall events.

Moreover, this study differentiates itself from existing research by specifically addressing the challenge of predicting rainfall distribution in a tropical region with high climate variability. The model was trained and tested using monthly data from 27 meteorological stations, covering the period from January 1, 2020, to March 31, 2022. This data-driven approach allows us to refine the SVM model by incorporating both spatial and temporal dynamics, which previous studies may have treated separately or with less emphasis on the temporal aspect.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

By developing this enhanced classification prediction method, this research aims to make a significant contribution to supporting water resource management policies, agricultural planning, and disaster management in Java Island related to future rainfall distribution.

2. RESEARCH METHODOLOGY

2.1 Research Stages

In this research, a system was built that includes several stages aimed at analyzing rainfall predictions. The following is the system flowchart design shown in Figure 1.

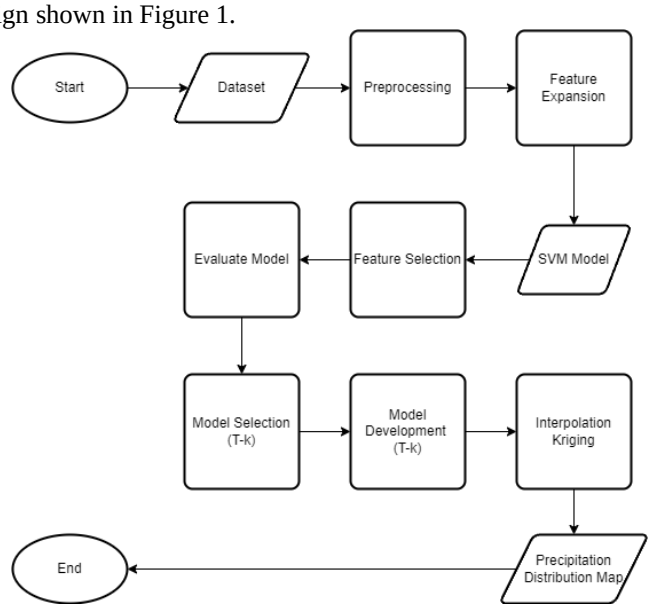


Figure 1. System Flowchart Design

Based on Figure 1, the flowchart depicts the system design for a project involving data preprocessing, feature expansion, and machine learning model training using SVM (Support Vector Machine). The process begins with the collection of the initial dataset rainfall data. The next stage is preprocessing, which includes fixing error datas, converting the dataset from daily to monthly format, normalizing the RR (Rainfall) values, and changing the RR labels into categorical data. After preprocessing, the monthly dataset is ready for further steps.

Feature expansion generates additional features, and oversampling is performed if there is an imbalance in the data. Important features are then selected to reduce dataset complexity, followed by splitting the data into training and testing sets. The SVM model undergoes hyperparameter tuning to optimize its performance. The training dataset is used to train the SVM model, which is subsequently tested using the testing dataset.

The model's performance is evaluated using appropriate metrics accuracy, precision, recall, and F1 values, and the best-performing SVM model is selected based on these results. This model is then used to predict the latest data. Kriging interpolation is applied to the predicted results to create a spatial prediction map, and the kriging results are visualized for better interpretation. The process concludes with the visualization of the results.

2.2 Dataset

The dataset was collected from various climatological stations across the island of Java. It includes input attributes such as Tn for minimum temperature (°C), Tx for maximum temperature (°C), Tavg for average temperature (°C), RH_avg for the duration of sunshine (hours), ss for average humidity (%), ff_x for maximum wind speed (m/s), ddd_x for wind direction at maximum speed (º), and ff_avg for average wind speed (m/s). The target attribute is RR, representing the rainfall rate (mm). Data ranged from January 2020 to March 2022 with detailed data per days. The initial dataset used in this study could be seen on Tabel 1.

Table 1. Dataset

Date	*Tn*	*Tx*	*Tavg*	*RH_avg*	*RR*	*ss*	*ff_x*	*ddd_x*	*ff_avg*	*Location*
01-01-2020	26	32.2	28.8	75	0	3.4	5	250	4	Geofisika_Tangerang
…	…	…	…	…	…	…	…	…	…	…
31-03-2022	22.6	33.5	26.7	77	0	7.2	5	320	1	Geofisika_Malang

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

2.3 Preprocessing

Data preprocessing is the process of converting raw data into a format that is easy to understand and use. Several steps are required during data preprocessing, as follows :

a. Fixing Data Error
In the preprocessing stage, fixing data errors is a critical step to ensure the dataset's quality and reliability for subsequent analysis. This process involves identifying and rectifying various data anomalies that could compromise the integrity of the analysis. Specifically, the procedure includes handling missing values and erroneous values, such as 9999 and 8888, which are often used as placeholders for missing or incorrect data. To address this, all missing values, as well as values of 9999 and 8888, are converted to 0. This transformation standardizes the dataset by eliminating placeholder values that could skew the results, thereby ensuring a cleaner and more consistent dataset for the following steps of feature expansion, selection, and model training.

b. Dataset change from Daily into Monthly
The process of changing the dataset from a daily format to a monthly format is a vital step in the preprocessing stage. This transformation involves aggregating the daily data into monthly data points, which can provide a more comprehensive view of trends and patterns over longer periods. By converting daily data into monthly data, the analysis can focus on broader trends rather than day-to-day fluctuations, which might be influenced by short-term anomalies or noise. This aggregation helps in smoothing out the data and making it more manageable and interpretable for subsequent analysis steps, such as feature expansion, selection, and model training. The monthly dataset thus serves as a more stable and reliable basis for building predictive models and deriving insights.

c. Normalize RR Value
After converting the dataset from daily to monthly format, normalizing the RR (Rainfall Rate) values is a crucial step in the preprocessing stage. This normalization process involves scaling the monthly RR values to a desired daily range, which enhances the comparability of the data and improves the performance of machine learning models. The normalization is carried out using the following mathematical formula :

$$Normalized_{Value_i} = Daily_{Lower} + \left(\frac{(value_i - Monthly_{Lower}) \times (Daily_{Upper} - Daily_{Lower})}{Monthly_{Upper} - Monthly_{Lower}}\right) \quad (1)$$

In this formula, $value_i$ represents the i-th monthly RR value that needs to be normalized. The terms $Monthy\_Lower$ and $Monthly\_Upper$ refer to the lower and upper bounds of the monthly RR values, respectively, while $Daily\_Lower$ and $Daily\_Upper$ denote the lower and upper bounds of the desired daily RR range. This transformation adjusts each monthly value to fit within the desired daily bounds by scaling it proportionally within the monthly bounds. It ensures that the RR values are standardized, making the data more suitable for analysis and modeling. Normalization helps in reducing the variability of the data and ensures that all features contribute equally to the analysis, leading to more accurate and reliable predictive models.

d. RR Label Change into Categorical
The process of changing RR (Rainfall Rate) labels into categorical values involves converting continuous RR values into categorical labels that represent different rainfall intensities. The categorization is based on the following rules:

Table 2. RR Value to Category Data

RR Value Range	*Categorical Label*	*Numerical Label*
≤ 0 mm	Cloudy	0
0 < RR < 20 mm	Light Rain	1
20 ≤ RR < 50 mm	Moderate Rain	2
50 ≤ RR < 100 mm	Heavy Rain	3
100 ≤ RR < 150 mm	Very Heavy Rain	4
≥ 150 mm	Extreme Rain	5

Table 2 summarizes the rules used to categorize RR values into distinct labels, facilitating easier analysis and interpretation of rainfall data by grouping continuous values into meaningful categories.

2.4 Feature Expansion

Feature expansion in this project involves utilizing time-based feature expansions in time-series analysis. Time-based feature expansions refer to incorporating previous time-step values of variables into the dataset as additional features [17]. For instance, for a variable at time t, the time-based feature expansion t-1 includes the value of the variable at t-1 as an additional feature. This approach allows the model to access crucial historical information necessary for making continuous predictions based on time-series data. As shown in Table 3, time-based expansion features have been systematically included to enhance the model's ability to learn from past trends.

In time series scenarios, we can increase the window size to include more lagged features. A common challenge data scientists face before expanding time-based features is determining the optimal window size [18]. A practical approach is to create a series of windows with varying widths and iteratively add and remove them from the dataset

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

to identify which configuration most positively impacts model performance. This iterative method aids in optimizing feature selection and improving the model's predictive accuracy in time-series forecasting tasks.

Table 3. Data Time-Based Expansion Feature

Time-based feature expansions Scenario	Combination	Feature Periode	Target
T-2	*1*	January 2020 – February 2020	March 2020
	2	February 2020 - March 2020	April 2020
	...	…	…
	25	January 2022 - February 2022	March 2022
T-3	*1*	January 2020 – March 2020	April 2020
	2	February 2020 - April 2020	May 2020
	...	…	…
	24	December 2021 - February 2022	March 2022
….		...	
T-24	*1*	January 2020 – December 2021	January 2022
	2	February 2020 - January 2022	February 2022
	3	March 2020 - February 2022	March 2022
T-25	*1*	January 2020 – January 2022	February 2022
	2	February 2020 - February 2022	March 2022

2.5 Oversampling

Oversampling is a data processing technique that increases the amount of data (samples) required to balance minority and majority classes. This helps improve model performance but can cause overfitting problems. The most widely used oversampling method is SMOTE (Synthetic Minority Over-sampling Technique) [19]. Sample synthesis is based on existing minority samples in the dataset, which can effectively balance minority classes with SMOTE. This is done without needing to replicate existing sample data but by synthetically creating additional instances. This process produces new samples in minority classes to achieve a better balance.

2.6 Feature Selection

The feature selection process in this project is conducted using *sklearn.feature_selection.SelectKBest* . This tool allows selecting the best features from the dataset based on scores calculated using various statistical metrics such as ANOVA for categorical features or chi-square for labeled categorical features. By choosing the most significant features, the complexity of the dataset can be reduced, which in turn can improve model performance by focusing on the most informative and relevant features for prediction.

2.7 Splitting Data

Splitting Data is the process of dividing a dataset into several parts for training and testing a model. The goal is to ensure that the model can be tested correctly and its performance can be measured after training. The data are divided into two parts: test data and training data. Test data are used to evaluate the model's performance and ensure accurate predictions. On the other hand, training data are used to train the model and refine it, allowing for accurate performance estimation on new, unseen data.

2.8 Support Vectore Machine

Support Vector Machine (SVM) is a supervised learning algorithm widely used for classification and regression tasks. SVM aims to find an optimal hyperplane in a high-dimensional space that separates classes of data points as widely as possible. This hyperplane is determined by support vectors, which are the data points closest to the hyperplane and influence its position and orientation [20].

SVM can efficiently handle nonlinear decision boundaries through the kernel trick, which implicitly maps the input features into a higher-dimensional feature space where a linear separation of classes is possible. Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels. Recent research has focused on enhancing SVM's performance and applicability in various domains. Studies have explored novel kernel functions tailored to specific data distributions to improve classification accuracy. Additionally, advancements in optimization techniques have addressed scalability issues, making SVM suitable for large-scale datasets [21]. SVM continues to

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

be a popular choice in machine learning due to its robustness, effectiveness in high-dimensional spaces, and ability to handle both linear and nonlinear classification tasks.

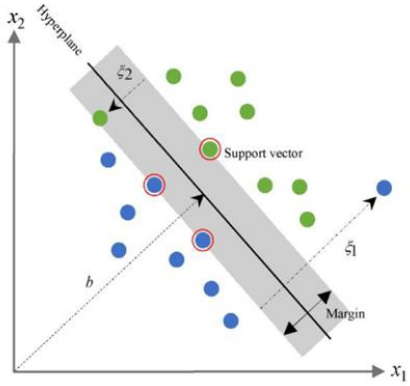


Figure 2. Illustration of support vector machine (SVM) to generalize the optimal separating hyperplane in linear separable data [22]

As illustrated in Figure 2, a hyperplane separates two classes of data. The hyperplane, also known as the decision boundary, is a geometric construct used to delineate one class from another. It helps determine the optimal separation between classes. Support vectors are the data points that lie closest to the hyperplane and define the margin. The margin is the distance between the hyperplane and the nearest support vector. The equation of the hyperplane in SVM is given by:

$$f(x) = w^T x + b \tag{2}$$

Where denotes the weight vector, represents the feature vector of the training data, and is the bias term, a scalar value. In SVM, the kernel function plays a critical role in transforming data from a low-dimensional space to a higher-dimensional space, facilitating better separation of classes. Several kernel functions are commonly used in SVM:

a. Linear Kernel
The linear kernel is the simplest kernel and is utilized when the data can be separated linearly. This kernel results in a hyperplane that is relatively easy to achieve compared to other kernels. Data that is linearly separable often exhibits high homogeneity and can be described as low-dimensional. The linear kernel function is expressed as:

$$K(x, y) = x^T y \tag{3}$$

Where and are data vectors. The linear kernel operates by evaluating the training data points, x_i , with binary class labels using the following conditions:

$$H \begin{cases} w^T x_i + b \geq +1 & \text{for } y_i = +1 \\ w^T x_i + b \leq -1 & \text{for } y_i = -1 \end{cases} \tag{4}$$

b. Polynomial Kernel
The polynomial kernel is used to address nonlinear data by transforming it into a higher-dimensional space using polynomial functions of a specified degree. This kernel is suitable for data with complex, non-linear structures. The polynomial kernel function is given by:

$$K(x, y) = (x^T y + C)^d \tag{5}$$

Where denotes the degree of the polynomial, is a constant parameter, and are feature vectors. This kernel is effective for capturing nonlinear relationships in the data.

c. Radial Basis Function (RBF) Kernel
The RBF kernel, or Gaussian kernel, is employed when there is limited prior knowledge about the data distribution. It measures the similarity between data points using a Gaussian function. The RBF kernel function is:

$$K(x, x') = exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{6}$$

Where and are data vectors, and is a parameter that controls the width of the Gaussian function. The RBF kernel is highly versatile and effective for many applications, especially with complex data patterns.

d. Sigmoid Kernel
The sigmoid kernel, while generally not as effective as the RBF kernel, can still be optimized to achieve comparable performance. The sigmoid kernel function is expressed as:

$$K(x, y) = \tanh(\alpha x^T y + c) \tag{7}$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

Where is a scaling parameter, is a bias term, and and are data vectors. The sigmoid kernel is often used in scenarios where the decision boundary is nonlinear in a higher-dimensional space.

2.9 Interpolation Kriging

Interpolation Kriging is a statistical method used to predict values at unsampled locations in space based on known measurement points. This method is commonly applied in spatial modeling and geo-statistics to generate optimal estimates and their variances. Kriging interpolation considers spatial dependence among data points by assigning different weights to nearby points in the prediction calculations. This approach allows for more accurate estimation and produces smooth prediction maps, thereby reducing interpolation errors typical in conventional linear methods [23].

In the context of rainfall prediction, Kriging interpolation is used to create spatial maps that depict the distribution of rainfall across a large area based on measured rainfall data at specific points. This method enables researchers or practitioners to better understand spatial patterns of rainfall, crucial for natural resource planning, environmental management, and disaster risk mitigation.

Kriging has several variants depending on the spatial correlation model used and assumptions about the spatial structure of the data. These methods, such as Ordinary Kriging, Universal Kriging, and others like Block Kriging, can be tailored to the data characteristics and analytical objectives. Choosing the appropriate model can impact prediction accuracy and the precision of result interpretation [24]. Kriging interpolation remains a popular approach in spatial analysis and geo-statistics due to its ability to address spatial heterogeneity issues by effectively utilizing local information.

In the research presented, Spherical Ordinary Kriging is the specific method employed. This variant of Kriging uses a spherical model to define the spatial correlation between data points. The spherical model assumes that the spatial correlation decreases with distance in a way that resembles a spherical shape, where the correlation is constant within a certain range and then levels off. By applying Spherical Ordinary Kriging, the study aims to leverage this model to achieve accurate rainfall predictions and generate reliable spatial maps. This method's choice is driven by its effectiveness in capturing the spatial structure of the rainfall data and providing precise estimates for areas with limited sampling points.

3. RESULT AND DISCUSSION

This section discusses the experimental results of the models created. The research constructs test scenarios for each existing model to analyze the performance and effectiveness of each model. The method employed in this study uses SVM modeling. Further explanations will be presented in the subsequent sections on the test results and their analysis.

3.1 Result

In this subsection, following the predetermined modeling stages, the subsequent step is evaluation. The purpose of evaluation is to assess the performance of SVM algorithm models using designated test data ratios. The results are analyzed using a comprehensive evaluation matrix comprising metrics such as accuracy, F1-score, precision, and recall. These metrics provide insights into the models' effectiveness across different test scenarios and datasets.

Evaluation conclusions are drawn based on weighted average calculations of metrics including F1-score, recall, accuracy, and precision. This approach offers a deeper understanding of the models' performance levels under various testing conditions and across diverse datasets.

a. Best T-K

Table 4. Best T-K SVM Time-Based Model

Lag Expansion	Parameters			Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Feature Count
	C	Gamma	Kernel					
T-2	100	0,8	rbf	88,889	92,593	88,889	89,206	7
T-3	100	0,3	rbf	88,889	92,593	88,889	89,206	9
T-4	10	0,8	sigmoid	90	92,5	90	89,048	5
T-5	10	0,8	sigmoid	90	92,5	90	89,048	5
T-6	10	0,5	sigmoid	90	92,5	90	89,048	8
T-7	10	0,4	sigmoid	90	92,5	90	89,048	10
T-8	10	0,3	sigmoid	90	93,333	90	89,333	11
T-9	10	0,7	sigmoid	90	93,333	90	89,333	5
T-10	10	0,6	sigmoid	90	93,333	90	89,333	6
T-11	10	0,6	sigmoid	90	93,333	90	89,333	6
T-12	10	0,2	sigmoid	90	93,333	90	89,333	15
T-13	10	0,2	sigmoid	90	93,333	90	89,333	17
T-14	10	0,2	sigmoid	90	93,333	90	89,333	17

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

T-15	10	0,2	sigmoid	90	93,333	90	89,333	17
T-16	100	0,5	sigmoid	87,5	91,667	87,5	87,5	16
T-17	100	0,5	sigmoid	87,5	91,667	87,5	87,5	16
T-18	100	0,7	poly	87,5	91,667	87,5	87,5	4
T-19	10	0,1	sigmoid	90	93,333	90	89,333	27
T-20	10	0,1	sigmoid	90	93,333	90	89,333	27
T-21	10	0,1	sigmoid	90	93,333	90	89,333	27
T-22	0,1	0,3	poly	87,5	79,167	87,5	82,5	27
T-23	10	0,1	sigmoid	90	93,333	90	89,333	26
T-24	0,1	0,3	poly	87,5	79,167	87,5	82,5	28
T-25	0,1	0,3	poly	87,5	79,167	87,5	82,5	29

As presented in Table 4, the comparative performance analysis of SVM models for rainfall rate prediction on an oversampled dataset across different data. The SVM model with the highest accuracy, achieving 90.0%, used time-based feature expansions (T-9) with parameters C=10, Gamma=0.7, Kernel=sigmoid, and Feature selection count = 5 feature. This configuration also resulted in an F1-score of 89.3%, precision of 93.3%, and recall of 90.0%. In contrast, the SVM model performed best with different time-based feature expansions (T-8) at parameters C=10, Gamma=0.3, and Kernel=sigmoid and Feature selection count = 11 feature, achieving an accuracy of 90.0%, an F1-score of 89.3%, precision of 93.3%, and recall of 90.0%. These results demonstrate the effectiveness of SVM models with time-based feature expansions in rainfall rate prediction, particularly with specific parameter settings and feature configurations. Based on the test results obtained, it is concluded that the use of time-based feature expansion expansions T-9 and T-8 yielded the best performance with identical accuracy, precision, recall, and F1-score values of 90%, 93.33%, 90%, and 89.33%, respectively. However, concerning feature count, T-9 outperformed T-8 as T-9 utilized only 5 features compared to 11 features in T-8. This difference resulted in T-9 having a smaller data dimensionality than T-8.

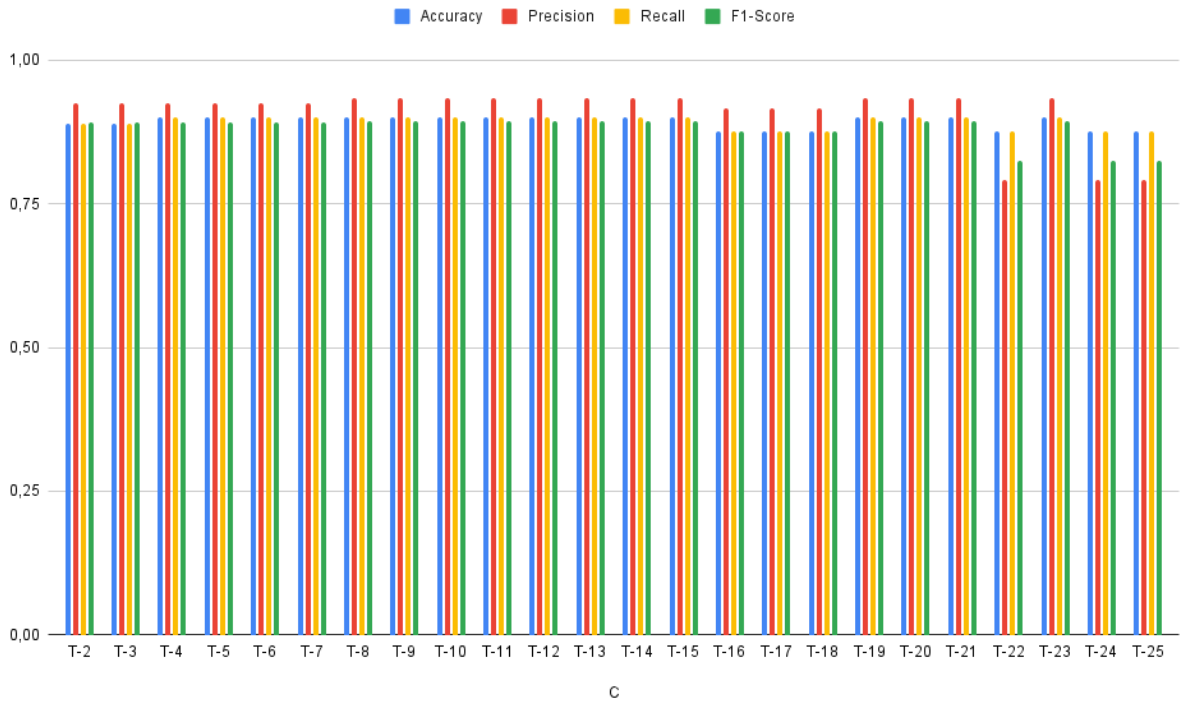


Figure 3. Best T-K SVM Time-Based Model

The results, illustrated in Figure 3, underscore the advantage of using time-based feature expansion T−9 for optimizing SVM model performance in rainfall rate prediction, with the selected features being with SVM parameters C=10, Gamma=0.7, Kernel=sigmoid, employing the following features: [RH_avg T-6, ff_avg T-1, ff_avg T-6, ff_avg T-9, ff_x T-4].

b. Prediction Result of T+K Model

The prediction results for rainfall on the latest date using the best time-based SVM model combined with Kriging interpolation are presented in the following tables. The model's performance is evaluated across various prediction intervals, from T+2 up to T+25, where each interval represents a day beyond the initial prediction date. The

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

predictions classify the expected rainfall intensity into different categories based on the RR value ranges, as outlined in Table 2.

Table 5 shows the predictions from T+2 to T+13, while Table 6 extends these predictions from T+14 to T+25. The numerical labels represent categorical rainfall predictions, where 0 indicates "Cloudy," 1 indicates "Light Rain," 2 indicates "Moderate Rain," 3 indicates "Heavy Rain," 4 indicates "Very Heavy Rain," and 5 indicates "Extreme Rain.".

Table 5. Best T+K SVM Time-Based Model for T+2 until T+13

Location	T+2	T+3	T+4	T+5	T+6	T+7	T+8	T+9	T+10	T+11	T+12	T+13
Geofisika_Tangerang	3	3	2	3	2	3	1	2	4	1	2	2
Klimatologi_Tangerang_Selatan	3	3	2	2	1	1	1	2	4	1	2	4
Meteorologi_Budiarto	3	4	3	3	3	3	3	3	3	3	3	3
Meteorologi_Maritim_Serang	4	1	3	3	1	1	1	2	4	1	2	4
Meteorologi_Soekarno_Hatta	3	3	3	3	3	3	3	5	3	3	3	3
Geofisika_Sleman	3	3	2	2	2	3	1	2	2	1	2	2
Meteorologi_Kemayoran	3	3	2	2	2	3	3	2	4	1	3	3
Meteorologi_Maritim_Tanjung_Priok	3	3	3	3	2	1	1	2	2	2	2	2
Geofisika_Bandung	2	2	3	3	1	1	1	2	4	1	2	4
Klimatologi_Bogor	3	2	2	2	2	3	1	2	4	1	2	2
Meteorologi_Citeko	4	4	3	3	3	3	3	5	5	5	2	2
Meteorologi_Kertajati	3	3	3	3	2	1	1	3	3	3	2	4
Meteorologi_Tegal	3	3	3	3	2	3	3	2	2	2	3	3
Meteorologi_Tunggul_Wulung	3	2	2	2	2	3	3	2	2	1	2	2
Meteorologi_Maritim_Tanjung_Emas	3	3	3	3	2	3	3	3	3	3	3	3
Meteorologi_Ahmad_Yani	5	3	3	3	3	2	2	3	3	3	3	3
Klimatologi_Semarang	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Sangkapura	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Perak_I	3	3	2	2	1	1	1	2	4	1	2	4
Meteorologi_Maritim_Tanjung_Perak	3	3	3	3	3	2	2	3	3	3	3	3
Meteorologi_Kalianget	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Juanda	3	3	3	3	3	2	2	3	3	3	3	3
Meteorologi_Banyuwangi	4	1	3	3	2	3	1	2	2	2	2	2
Klimatologi_Malang	4	4	2	2	2	3	3	3	3	3	2	3
Geofisika_Pasuruan	3	4	5	5	5	2	1	3	3	3	2	4
Geofisika_Nganjuk	2	2	3	3	1	1	1	2	2	1	2	4
Geofisika_Malang	3	3	2	2	2	3	1	2	2	1	2	4

Table 6. Best T+K SVM Time-Based Model for T+14 until T+25

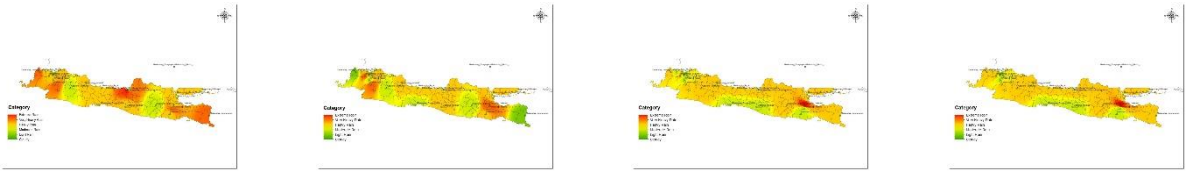
Location	T+14	T+15	T+16	T+17	T+18	T+19	T+20	T+21	T+22	T+23	T+24	T+25
Geofisika_Tangerang	2	2	3	3	3	3	3	3	3	3	3	3
Klimatologi_Tangerang_Selatan	4	1	3	3	3	2	2	4	2	2	2	2
Meteorologi_Budiarto	3	3	2	2	4	3	3	3	3	3	3	3

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

Meteorologi_Maritim_Serang	4	1	2	2	3	1	1	1	1	1	2	2
Meteorologi_Soekarno_Hatta	3	3	3	3	3	3	3	3	3	3	3	3
Geofisika_Sleman	2	2	3	3	3	3	3	1	2	3	2	2
Meteorologi_Kemayoran	3	3	3	3	3	3	3	3	3	3	2	3
Meteorologi_Maritim_Tanjung_Priok	2	2	3	3	3	3	3	3	3	3	3	3
Geofisika_Bandung	2	2	2	2	3	2	2	2	2	3	2	2
Klimatologi_Bogor	2	2	2	2	3	2	2	3	2	3	2	2
Meteorologi_Citeko	2	2	3	3	3	2	3	3	3	3	3	2
Meteorologi_Kertajati	4	1	3	3	3	1	1	1	2	1	2	2
Meteorologi_Tegal	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Tunggul_Wulung	2	2	2	3	3	3	3	3	2	3	2	2
Meteorologi_Maritim_Tanjung_Emas	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Ahmad_Yani	3	3	3	3	3	3	3	3	3	3	3	3
Klimatologi_Semarang	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Sangkapura	3	3	3	3	2	3	3	3	3	3	3	3
Meteorologi_Perak_I	2	2	3	3	3	2	2	2	2	3	2	3
Meteorologi_Maritim_Tanjung_Perak	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Kalianget	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Juanda	3	3	3	3	3	3	3	3	3	3	3	3
Meteorologi_Banyuwangi	2	2	2	2	1	2	2	3	2	3	2	2
Klimatologi_Malang	3	3	2	1	4	3	3	3	3	3	3	3
Geofisika_Pasuruan	4	1	2	4	3	1	1	1	1	1	1	2
Geofisika_Nganjuk	2	2	3	3	3	2	3	3	2	3	2	2
Geofisika_Malang	2	2	3	3	3	2	2	3	2	3	2	2

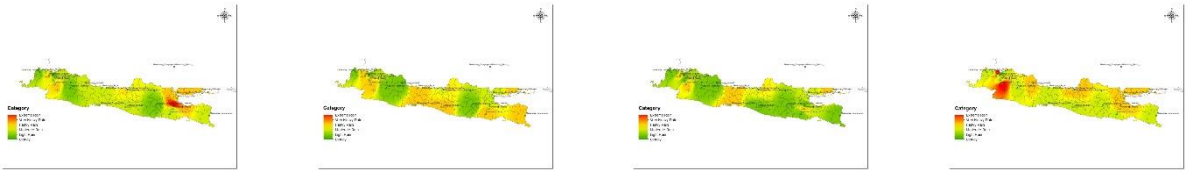
c. Visualization of Rainfall Classification Results

After predicting rainfall rates using the SVM model with time-based feature expansion T-9, the results were interpolated using Kriging interpolation to create a spatial prediction map. Kriging is a geostatistical interpolation method that estimates values at unsampled locations based on the spatial correlation of the sampled data points.



(a) (b) (c) (d)

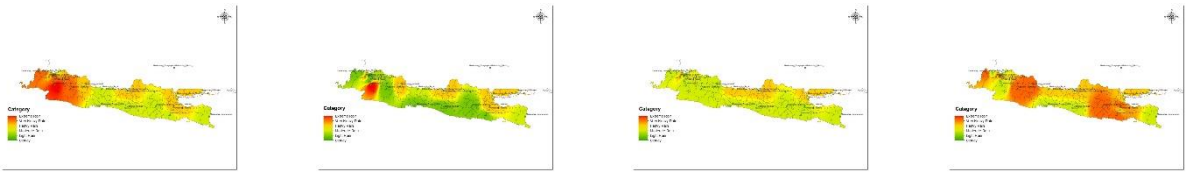
Figure 4. Prediction Map of Rainfall Distribution of May 2022 to August 2022



(a) (b) (c) (d)

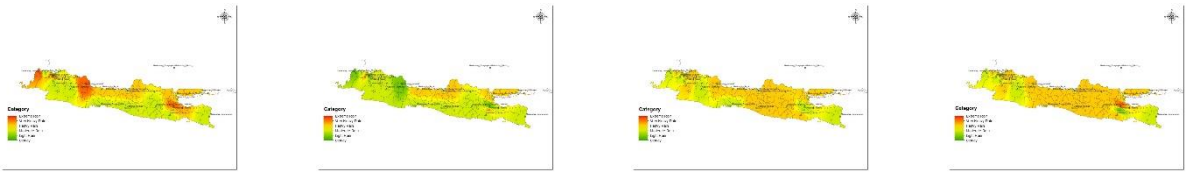
Figure 5. Prediction Map of Rainfall Distribution of September 2022 to December 2022

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758



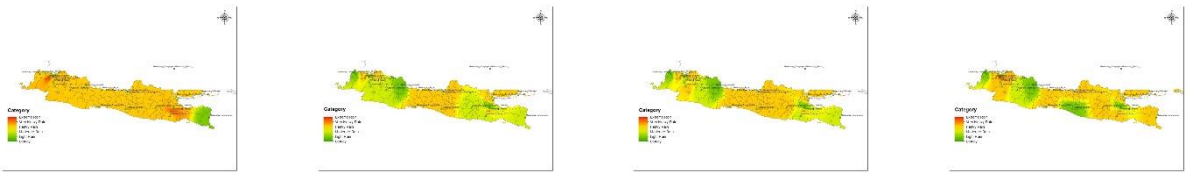
(a) (b) (c) (d)

Figure 6. Prediction Map of Rainfall Distribution of January 2023 to April 2023



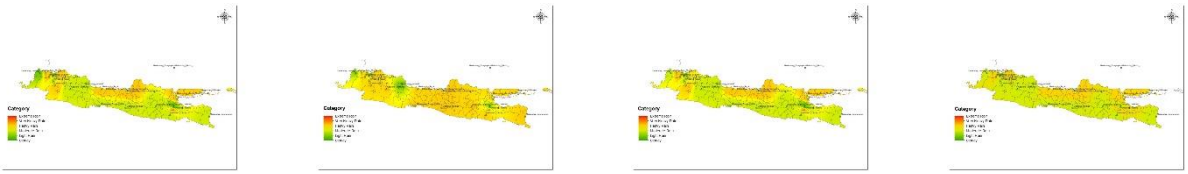
(a) (b) (c) (d)

Figure 7. Prediction Map of Rainfall Distribution of May 2023 to August 2023



(a) (b) (c) (d)

Figure 8. Prediction Map of Rainfall Distribution of September 2023 to December 2023



(a) (b) (c) (d)

Figure 9. Prediction Map of Rainfall Distribution of January 2024 to April 2024

The visualizations of Kriging interpolation results in Figures 4 to 9 reveal the spatial distribution of predicted rainfall rates across Java Island for different periods. These maps use varying colors to represent different rainfall intensities, making it easier to identify regions expected to experience moderate, heavy, or extreme rainfall. Figure 4 illustrates the predicted rainfall distribution from May 2022 to August 2022. This period shows a pattern of heavier rainfall predominantly in the western parts of Java, particularly around the Banten and West Java regions, while the eastern parts tend to exhibit moderate rainfall. Figure 5 shows the rainfall predictions from September 2022 to December 2022. The map indicates a shift, where moderate to heavy rainfall begins to spread towards central Java, while western regions maintain consistent heavy rainfall. Figure 6 presents the predictions for January 2023 to April 2023. During these months, the intensity of rainfall appears to be more evenly distributed across the island, with both western and eastern regions experiencing significant rainfall events. Figure 7 depicts the predicted rainfall from May 2023 to August 2023. Similar to the previous year, western Java again shows heavy rainfall, indicating a recurring seasonal pattern. The eastern regions continue to experience less intense but still significant rainfall. Figure 8 focuses on predictions from September 2023 to December 2023. The central and eastern regions of Java show an increase in rainfall intensity, with some areas moving towards heavier categories, while western parts continue their pattern of consistent rainfall. Figure 9 covers the predictions for January 2024 to April 2024. The map shows a return to a more balanced distribution of rainfall, with moderate to heavy rainfall spreading across much of Java Island.

The consistency of these spatial patterns over multiple periods, as seen in Figures 4 to 9, demonstrates that the Kriging model effectively captures the temporal and spatial dynamics of rainfall distribution. This stable pattern highlights the model's utility in predicting seasonal rainfall trends, which is crucial for informing time-based feature expansion in SVM modeling. By linking each explanation directly to its corresponding figure, the analysis provides a clearer understanding of how rainfall patterns evolve over time and space. This enhanced clarity supports more informed decision-making in agriculture, water resource management, and disaster preparedness. Additionally, the observation of consistent seasonal patterns, particularly the recurring heavy rainfall in western

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

Java, underlines the importance of selecting appropriate features such as temperature, humidity, and air pressure to improve predictive model accuracy.

3.2 Discussion

The experimental results presented highlight the effectiveness of Support Vector Machine (SVM) models with time-based feature expansions in predicting rainfall rates, particularly with the use of Spherical Ordinary Kriging for spatial interpolation. This discussion provides a deeper analysis of the results and their implications.

The analysis reveals that the SVM models with time-based feature expansions T-9 and T-8 yielded the highest performance metrics, including accuracy, precision, recall, and F1-score, all at 90.0%, 93.3%, 90.0%, and 89.3%, respectively. Both configurations demonstrated robust performance across the evaluated metrics, indicating that the SVM models are well-suited for rainfall prediction tasks. However, T-9, which uses only 5 features, outperforms T-8, which uses 11 features, in terms of model efficiency and simplicity. This suggests that a more concise feature set can lead to equally effective or even better performance while potentially reducing computational complexity and risk of overfitting.

The choice of parameters, such as C, Gamma, and Kernel type, significantly influences the SVM model's performance. For instance, the SVM model with parameters C=10, Gamma=0.7, and Kernel=sigmoid (T-9) provided a high degree of accuracy and precision with fewer features. This result underscores the importance of optimizing SVM parameters and feature selection to achieve high predictive accuracy. The sigmoid kernel, in particular, appears to be effective in handling the non-linear relationships present in rainfall data.

The integration of Kriging interpolation with SVM predictions to generate spatial maps provides valuable insights into the spatial distribution of rainfall rates. The Kriging interpolation method, specifically Spherical Ordinary Kriging, was used to estimate values at unsampled locations, creating detailed spatial prediction maps. These maps reveal areas with varying rainfall intensities and support a nuanced understanding of spatial patterns. This is critical for applications in resource management and disaster preparedness, as it allows for targeted interventions and better planning.

The results suggest that the SVM models with time-based feature expansions are highly competitive compared to other methods. While the exact performance metrics of alternative models were not provided in this analysis, the consistent high performance of the SVM models with time-based expansions indicates their robustness. Future research could involve comparing these SVM models with other machine learning techniques or hybrid approaches to further validate their effectiveness and identify potential improvements.

The ability to accurately predict rainfall distribution has significant practical implications. In agriculture, it helps in planning irrigation schedules and managing crop yields. For water resource management, it assists in forecasting water availability and planning reservoir operations. Moreover, accurate rainfall predictions are crucial for disaster risk reduction, enabling timely warnings and better preparedness for extreme weather events.

4. CONCLUSION

This research demonstrates the effectiveness of employing time-based feature expansions within Support Vector Machine (SVM) models for predicting rainfall rates, using historical data to forecast future weather patterns. Among the evaluated time-based feature expansions, the T-9 model achieved superior performance, with an accuracy of 90.00%, an F1-score of 89.33%, a precision of 93.33%, and a recall of 90.00%, while utilizing only 5 features. This indicates that the T-9 model consistently outperforms other configurations, showcasing the importance of efficient feature selection and parameter optimization. The study highlights the critical role of feature engineering and model parameter optimization in enhancing predictive accuracy and efficiency. The successful integration of SVM models with time-based feature expansions and Spherical Ordinary Kriging interpolation not only underscores their high performance but also demonstrates their potential for practical applications. The spatial maps generated through Kriging interpolation provide valuable insights into rainfall distribution, supporting applications in agriculture, water resource management, and disaster preparedness. Future research should focus on exploring additional feature expansion techniques, kernel functions, and model configurations to further refine and enhance prediction models. Advancements in these areas are essential for improving weather forecasting accuracy and addressing real-world challenges in meteorology and environmental management. By continuing to innovate and optimize these models, we can better support decision-making processes and resource management across various domains.

REFERENCES

[1] A. H. Syafrina, M. D. Zalina, and L. Juneng, “Spatial and temporal characteristics of rainfall trends over Sumatra Island, Indonesia,” Theor Appl Climatol, vol. 141, no. 1–2, pp. 163–173, 2020.
[2] W. Supari and E. Yulihastin, “The influence of atmospheric circulation on extreme rainfall events over Indonesia,” Atmos Res, vol. 218, pp. 156–162, 2019.
[3] T. Ferijal, O. Batelaan, and M. Shanafield, “Spatial and temporal variation in rainy season droughts in the Indonesian Maritime Continent,” Journal of Hydrology, vol. 603, no. 126999, 2021.
[4] S. Ahmad, A. Kalra, and H. Stephen, “A comparison of ensemble and hybrid SVM techniques for rainfall prediction,” Water Resources Management, vol. 34, no. 11, pp. 3535–3551, 2020.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024 Page: 1071−1082
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5758

[5] L. Zhao, H. Du, and Y. Han, "Support vector machine-based rainfall forecasting," Environmental Modelling & Software, vol. 144, no. 105123, 2021.
[6] Y. Zhang, J. Wang, and X. Li, "Application of Support Vector Machine in Rainfall Prediction: A Case Study in a Tropical Region," J Hydrol (Amst), vol. 596, no. 126042, 2022.
[7] H. Liu, X. Zhang, and J. Wang, " Improving rainfall prediction accuracy using hybrid SVM-Kriging model," J Hydrol (Amst), vol. 598, no. 126413, 2021.
[8] S. Kumar, R. Gupta, and V. Sharma, "Enhancing Rainfall Prediction Accuracy Using SVM and Kriging in Tropical Climates," Water Resour Res, vol. 59, no. 5, 2023.
[9] Z. Ahmed and M. K. Faisal, "Advancements in Support Vector Machines for Environmental Modeling," Journal of Environmental Informatics, vol. 47, no. 1, pp. 40–55, 2021, doi: 10.3808/jei.2021.47.1.40.
[10] D. Luo, X. Wang, and L. Yang, "Improving Spatial Rainfall Estimation with Enhanced Kriging Techniques," Hydrology Research, vol. 54, no. 2, pp. 345–361, 2022, doi: 10.2166/nh.2022.065.
[11] E. Zhao and P. Tan, "Integration of Machine Learning and Kriging for Spatial Data Analysis: A Review," Comput Geosci, vol. 151, pp. 104–117, 2021, doi: 10.1016/j.cageo.2021.104117.
[12] Y. Wang, M. Xu, and X. Zhao, "Comparative analysis of kriging interpolation and support vector machine regression for spatial data," Environ Earth Sci, vol. 78, no. 3, p. 81, 2019.
[13] J. Park, J. Kim, and S. Lee, "Application of machine learning techniques for flood prediction in urban areas: A review," Water (Basel), vol. 12, no. 1, p. 352, 2020.
[14] N. Ali, M. B. Rahman, and H. U. Ahmed, "Hybrid Models of SVM and Kriging for Improved Spatial Predictions," Environ Monit Assess, vol. 196, no. 9, pp. 1–15, 2024, doi: 10.1007/s10661-024-11628-0.
[15] J. Sun, T. Wu, and Q. Li, "Recent Developments in Kriging Methods for Accurate Spatial Data Modeling," Spat Stat, vol. 40, p. 100546, 2024, doi: 10.1016/j.spasta.2024.100546.
[16] L. Cheng, Y. Zhang, and Y. Wu, "Predicting extreme weather events using hybrid machine learning models," J Environ Manage, vol. 325, no. 116401, 2023.
[17] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018.
[18] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.
[19] K. Shin, J. Han, and S. Kang, "MI-MOTE: Multiple imputation-based minority oversampling technique for imbalanced and incomplete data classification," Inf Sci (N Y), vol. 575, pp. 80–89, 2021.
[20] A. Smith and B. Johnson, "Efficient parameter tuning for support vector machines in large-scale datasets," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 8, pp. 2404–2415, 2019.
[21] J. Lee, H. Park, and S. Kim, "Enhanced support vector machines using adaptive kernel functions," Pattern Recognit Lett, vol. 131, pp. 123–130, 2020.
[22] K. Ousmane et al., "Novel Classification Method of Spikes Morphology in EEG Signal Using Machine Learning," Procedia Comput Sci, vol. 148, pp. 70–79, Jul. 2019, doi: 10.1016/j.procs.2019.01.010.
[23] P. Singh and P. Verma, "A comparative study of spatial interpolation technique (IDW and Kriging) for determining groundwater quality," GIS and geostatistical techniques for groundwater science, pp. 43–56, 2019.
[24] M. P. Lucas et al., "Optimizing automated kriging to improve spatial interpolation of monthly rainfall over complex terrain," J Hydrometeorol, vol. 23, no. 4, pp. 561–572, 2022.