VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i2.6665

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

# FORECASTING UPWELLING IN LAKE MANINJAU USING VECTOR AUTOREGRESSIVE, SUPPORT VECTOR MACHINE AND DASHBOARD VISUALIZATION

**Fakhrus Syakir[1]; Muhammad Irhamsyah[1*]; Melinda Melinda[1]; Yunidar Yunidar[1]; Zuhelmi Zuhelmi[1]; Rizka Miftahujjannah[1]**

Department of Electrical Engineering and Computer[1]
Universitas Syiah Kuala, Banda Aceh, Indonesia[1]
www.usk.ac.id[1]
fakrus@mhs.unsyiah.ac.id, irham.ee@usk.ac.id*, melinda@usk.ac.id, yunidar@usk.ac.id, zulhelmi.za@usk.ac.id, rizkamiftahujjannah03@gmail.com

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



***Abstract***— *Lake Maninjau experiences periodic upwelling events that disrupt water quality, harm fish stocks, and pose socioeconomic challenges to surrounding communities. This study aimed to enhance upwelling prediction accuracy by integrating Vector Autoregressive (VAR) time series modelling with Support Vector Machine (SVM) classification. A five-year dataset (2020–2024) of daily climate variables surface temperature, precipitation, and wind speed was collected from NASA. Data stationarity was confirmed using Box-Cox transformations and Augmented Dickey-Fuller tests, while Granger Causality analysis revealed bidirectional relationships among the variables. The optimal forecasting model, VAR(17), was selected based on the Akaike Information Criterion (AIC), ensuring residuals met white-noise criteria. K-means clustering then labelled potential upwelling days, and these labels were employed to train SVM classifiers. An interactive dashboard was developed using Python and Streamlit to facilitate real-time forecasts and classification outputs. The VAR(17) model produced highly accurate forecasts, reflected by minimal error metrics (e.g., RMSE < 0.60). SVM classification of potential upwelling events achieved strong performance, consistently attaining F1-scores above 0.95. By merging time series forecasts with event classification, the hybrid VAR–SVM framework outperformed single-method approaches in identifying and predicting upwelling episodes. This integrated modelling strategy effectively addresses the complexity of upwelling in Lake Maninjau, enabling timely decision-making for fisheries management and local tourism stakeholders. Future work may incorporate additional environmental indicators (e.g., dissolved oxygen, pH) and extend dashboard functionalities to bolster sustainable resource management and community resilience.*

***Keywords***: *forecasting, lake maninjau, support vector machine, time series, upwelling.*

**Intisari**— *Danau Maninjau mengalami peristiwa upwelling secara berkala yang mengganggu kualitas air, merusak stok ikan, dan menimbulkan tantangan sosial ekonomi bagi masyarakat sekitar. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi upwelling dengan mengintegrasikan pemodelan deret waktu Vector Autoregressive (VAR) dengan klasifikasi Support Vector Machine (SVM). Dataset lima tahun (2020-2024) dari variabel iklim harian suhu permukaan, curah hujan, dan kecepatan angin dikumpulkan dari NASA. Stasioneritas data dikonfirmasi dengan menggunakan transformasi Box-Cox dan uji Augmented Dickey-Fuller, sementara analisis Kausalitas Granger menunjukkan hubungan dua arah di antara variabel-variabel tersebut. Model peramalan yang optimal, VAR (17), dipilih berdasarkan Akaike Information Criterion (AIC), yang memastikan residual memenuhi kriteria white-noise. Pengelompokan K-means kemudian memberi label pada hari-hari upwelling potensial, dan label-label ini digunakan untuk melatih pengklasifikasi SVM. Dasbor interaktif dikembangkan dengan menggunakan Python dan Streamlit untuk memfasilitasi prakiraan waktu nyata dan hasil klasifikasi. Model VAR (17) menghasilkan prakiraan yang sangat akurat, yang tercermin dari*

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6665

*metrik kesalahan yang minimal (misalnya, RMSE <0,60). Klasifikasi SVM untuk kejadian upwelling potensial mencapai kinerja yang kuat, secara konsisten mencapai nilai F1 di atas 0,95. Dengan menggabungkan prakiraan deret waktu dengan klasifikasi kejadian, kerangka kerja hibrida VAR-SVM mengungguli pendekatan metode tunggal dalam mengidentifikasi dan memprediksi episode upwelling. Strategi pemodelan terpadu ini secara efektif mengatasi kompleksitas upwelling di Danau Maninjau, sehingga memungkinkan pengambilan keputusan yang tepat waktu untuk pengelolaan perikanan dan pemangku kepentingan pariwisata lokal. Penelitian di masa depan dapat memasukkan indikator lingkungan tambahan (misalnya, oksigen terlarut, pH) dan memperluas fungsi dasbor untuk meningkatkan pengelolaan sumber daya yang berkelanjutan dan ketahanan masyarakat.*

***Kata Kunci**: peramalan, danau maninjau, support vector machine, deret waktu, umbalan.*

## INTRODUCTION

Upwelling is a critical oceanographic process where deep, nutrient-rich waters rise to the surface, significantly influencing marine ecosystems by altering nutrient distributions and promoting biological productivity[1]. Although commonly associated with ocean environments, upwelling can also occur in lacustrine systems, particularly in volcanic lakes such as Lake Maninjau in West Sumatra. In these lakes, upwelling is driven by environmental factors, including temperature gradients, wind patterns, and rainfall, impacting water quality and the distribution of aquatic organisms.

Lake Maninjau, formed within a volcanic caldera, presents a unique ecosystem where upwelling events have profound ecological and socioeconomic implications. These events can lead to sudden changes in water chemistry, such as increased turbidity and decreased oxygen levels, adversely affecting fish populations and biodiversity[2]. The lake's health is thus directly linked to the well-being of the surrounding communities, making the understanding and prediction of upwelling events a matter of significant importance.

The local communities around Lake Maninjau rely heavily on the lake for fisheries and tourism, both of which are sensitive to environmental changes induced by upwelling events[3]. Fisheries are a primary source of income, and fluctuations in fish stocks due to upwelling can have immediate economic consequences. Similarly, the lake's aesthetic appeal, crucial for tourism, can be diminished during upwelling events. Despite the critical need for accurate predictions to mitigate these impacts, existing forecasting methods face limitations in capturing the complex dynamics of upwelling in lake environments.

Current prediction methods predominantly employ either time series analysis or machine learning classification techniques independently. Time series models like the Vector Autoregressive (VAR) model are effective in modeling temporal dependencies within environmental data but cannot classify specific events such as upwelling[4]. Conversely, machine learning classifiers such as Support Vector Machines (SVM) excel in categorizing events but often disregard temporal relationships essential for accurate forecasting in dynamic systems. SVMs operate by identifying the optimal hyperplane that separates different classes in a high-dimensional space, making them particularly effective for binary classification tasks[5], [6]. Previous research has demonstrated the potential of combining these methods in other domains, yet their application to lake upwelling prediction remains unexplored.

The main research problem addressed in this study is the inadequacy of existing models to accurately predict upwelling events in Lake Maninjau due to their inability to simultaneously capture temporal dependencies and event-specific characteristics influenced by local environmental factors. These limitations hinder effective decision-making and proactive management of the lake's resources by the local community and stakeholders.

To overcome these challenges, we propose a novel hybrid modeling approach that integrates VAR time series analysis with SVM classification. This hybrid model aims to leverage the strengths of both methods: the VAR component models the temporal patterns in environmental data such as temperature, wind speed, and rainfall, while the SVM classifier identifies and categorizes upwelling events based on the patterns extracted by the VAR model. This combined approach is anticipated to enhance prediction accuracy and provide more reliable forecasts of upwelling events.

The specific implementation involves collecting key environmental data from reputable sources, including NASA and local monitoring stations. The data undergoes preprocessing to address any inconsistencies or gaps. The VAR model is then applied to uncover temporal dependencies, and its outputs serve as features for the SVM classifier. The SVM model classifies the events into

upwelling or non-upwelling categories, effectively capturing temporal and event-specific characteristics. This methodology builds upon previous research that has successfully utilized hybrid models in other environmental prediction contexts [7], [8].

To ensure that the predictive insights are accessible and beneficial to the local community, we have developed an interactive dashboard using Python and Streamlit. The dashboard provides forecasts, visualizations of historical trends, and user-friendly interfaces for data exploration. By making the model's outputs readily available, the dashboard empowers stakeholders to make informed decisions, enhancing the practical impact of the research. The contributions of this study are as follows:

1. Develop a hybrid VAR–SVM model to enhance the prediction accuracy of upwelling events in Lake Maninjau by integrating time series forecasting and event classification techniques.
2. Implement an optimal VAR model to capture temporal dependencies in climate data such as surface temperature, precipitation, and wind speed, ensuring robust and reliable forecasting.
3. Apply K-means clustering to label potential upwelling days, which are then used to train an SVM classifier that accurately identifies and classifies upwelling events.
4. Develop an interactive forecasting dashboard using Python and Streamlit, providing stakeholders with visualizations and decision-making support for local fisheries and tourism management.
5. Contribute to sustainable resource management by offering predictive tools that help mitigate the socioeconomic impacts of upwelling events on local communities.

Despite studies employing hybrid models for environmental forecasting, there is a notable scarcity of research applying such approaches to lake upwelling prediction. Previous studies have primarily focused on either time series analysis or machine learning classification in isolation, without integrating the two to capture the multifaceted nature of upwelling events [9], [10] This gap in the literature underscores the need for innovative models that can address temporal dynamics and event classification, particularly in lake ecosystems, where data characteristics may differ from oceanic environments.

This study aims to develop and validate a hybrid VAR-SVM model to improve the prediction of upwelling events in Lake Maninjau, thereby addressing a critical research gap. It is hypothesized that such a hybrid model will surpass conventional methods by accurately capturing both temporal dependencies and event-specific dynamics influenced by environmental factors. The novelty of this research lies not only in the hybrid modeling approach but also in the implementation of an interactive dashboard, bridging the gap between advanced predictive analytics and practical community applications. By enhancing the accuracy and accessibility of upwelling predictions, this study aims to promote sustainable resource management and contribute to the socioeconomic well-being of the Lake Maninjau community.

The predictive accuracy demonstrated by this study's hybrid VAR–SVM model aligns favorably with recent hybrid approaches utilizing VAR or SVM methodologies in climate and hydrological forecasting. In previous research, hybrid models integrating VAR have consistently exhibited low error metrics, such as in forecasting dam water levels, where VAR-based approaches achieved notably low RMSE and MAE values, demonstrating robustness in capturing linear climate-driven relationships [11]. Furthermore, integrating VAR with nonlinear machine learning methods like Artificial Neural Networks (ANN) or Support Vector Machines (SVM) has been shown to substantially reduce forecasting errors compared to single-method applications, highlighting the strength of hybrid modeling frameworks[11]. Specifically, semi-supervised SVM models applied to lake upwelling predictions reported exceptional classification performance, achieving an F1-score as high as 0.985 with approximately 99.5% precision, underscoring the effectiveness of SVM in accurately classifying complex environmental events [12].

Similarly, in drought forecasting contexts, hybrid approaches incorporating SVM alongside signal processing methods, such as wavelet transformations, have consistently outperformed standard statistical models (e.g., ARIMA), resulting in significantly improved RMSE values and overall predictive accuracy [13]. Recent literature emphasizes that hybrid VAR–SVM methodologies consistently deliver superior accuracy and reliability in forecasting critical environmental phenomena. Thus, the results presented in this study, characterized by low error metrics (MAE, RMSE) and strong classification performance (high F1-score), validate the efficacy and suitability of the hybrid VAR–SVM approach for precise and reliable lake upwelling prediction..


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

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6665

## MATERIALS AND METHODS

### Research Design and Timing

This study employed an observational design combined with a predictive modeling approach to analyze climate data from the Maninjau Lake region in Agam District, West Sumatra Province. The dataset covered a five-year period, from January 1, 2020, to December 31, 2024. The results of the analysis were integrated into an interactive dashboard designed to support informed decision-making among floating net cage fish farmers.

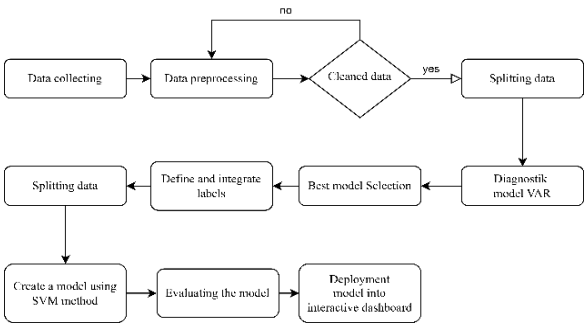
The climate data were obtained from the official and reliable website of the National Aeronautics and Space Administration (NASA), accessible at https://power.larc.nasa.gov/ [14]. The data acquisition process followed a series of steps. First, the POWER Data Access Viewer was accessed. The *Agroclimatology* data category was selected to align with the study's objectives. The temporal resolution was set to daily. Geographic coordinates were specified at a latitude of −0.399680 and a longitude of 100.200037. The selected time range spanned from January 1, 2020, to December 31, 2024. Three key climate parameters were then selected: surface temperature (°C), precipitation (mm), and mean wind speed at a 10-meter height (m/s), given their relevance to upwelling processes [7]. Finally, the data were exported in the comma-separated values (CSV) format to facilitate further processing. In total, the dataset comprised 1,827 daily observations, providing a comprehensive basis for climate pattern analysis in the study area.

### Software and Statistical Analysis

Data analysis utilized R software (version 4.4.1) and Python (version 3.12). In the initial phase, R was employed for comprehensive data preprocessing and advanced statistical processes, including data integration and cleaning. Descriptive statistics were derived by calculating the minimum, median, mean, and maximum values to understand central tendencies and by determining the interquartile range (IQR) and standard deviation to capture variability. Python was then used to create boxplots with jitter to visualize data distributions.

### Research stages

This study was structured into several sequential stages: data collection, data preprocessing, initial data splitting, diagnosing and selecting the best forecasting model, defining and integrating labels through clustering, secondary data splitting, building a classification model using Support Vector Machines (SVM), model evaluation, and deployment into an interactive dashboard. A brief methodology is presented in Figure 1.




Source : (Research Results, 2025)

Figure 1 Research Stages

Figure 1 shows the flow of the research; in the data collection phase, comprehensive climate data, including temperature, precipitation, and wind speed, were obtained from NASA's data access viewer. The subsequent model development had two primary objectives: forecasting climate indicators and classifying potential upwelling events. Initially, data underwent preprocessing, which included assessing data stationarity through the Box-Cox transformation and the Augmented Dickey-Fuller test. The Box-Cox transformation is expressed mathematically as:

$$T(Y_t) = \frac{{Y_t}^{\lambda} - 1}{\lambda} \quad (1)$$

Where T(Yt) is the transformation function applied to the data Y at time t, and λ is the transformation parameter. After transforming the data and verifying stationarity, a Granger causality analysis was performed to investigate relationships among the variables [12]. The dataset was then split into 80% training (876 rows) and 20% test data (219 rows), and time series methods were applied to construct forecasting models. Residuals from each model were examined for white noise properties and adherence to multivariate assumptions. A Vector Autoregression (VAR) model was selected for the time series forecasting. The VAR model can be represented by Equation (2):

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \alpha_t \quad (2)$$

Where Yt is a vector of endogenous variables at time T, Ai is the coefficient matrix for lag i, p is the order of the VAR model, and αt is the white-noise residual vector. A portmanteau test was applied to

confirm the model's validity, ensuring the residuals indeed resembled white noise.

Model selection was guided by minimizing the Akaike Information Criterion (AIC), which balances predictive accuracy and model simplicity. The AIC is derived from the model's likelihood and imposes a penalty based on the number of parameters, as shown in Equation (3):

$$AIC = -2\log L + 2k \quad (3)$$

where L is the likelihood and k denotes the number of parameters, guided the selection of the most appropriate VAR model by balancing model complexity against predictive accuracy [13], [15], [16].

After identifying the best-performing VAR model, unlabeled forecasted data were classified into two clusters ('potential upwelling' and 'no potential upwelling') through K-means clustering. This algorithm partitions datasets into clusters based on centroid proximity, iteratively minimizing intra-cluster variance [19], [20], [21]. Techniques such as k-means have been recognized for improving centroid initialization and overall clustering quality [17], [18].

The new labels generated from clustering were integrated into the dataset, which was again partitioned into training (80%) and testing (20%) subsets. Subsequently, a Support Vector Machine (SVM) classification model was developed. The SVM method, proposed by Cortes and Vapnik, identifies optimal hyperplanes that maximize margins between different data classes, efficiently handling both linear and nonlinear classification through kernel functions [24], [25], [26].

The forecasting component was evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Squared Error (MSE) [27]. The classification model's performance was assessed using accuracy, precision, recall, and F1-score metrics. High predictive accuracy and robust classification outcomes were achieved, with metrics frequently exceeding 90%.

Lastly, the integrated VAR–SVM solution was deployed in an interactive dashboard developed using R for forecasting climate indicators and Python for SVM-based classification. The dashboard enhances practical decision-making and operational responsiveness, underscoring the synergistic integration of statistical modeling and machine learning techniques consistent with methodologies employed across hydrological and climatological research domains[12], [13], [28].

## RESULTS AND DISCUSSION

### Descriptive Statistics Analysis

Descriptive statistics are a fundamental step in data analysis, and they offer a preliminary dataset overview through numerical summaries and visual representations. This approach facilitates the transformation of raw, often complex, data into a more interpretable form, allowing key characteristics to be conveyed concisely and efficiently. Through the use of central tendency measures (such as mean, median, and mode), dispersion metrics (such as standard deviation and range), and graphical tools (including histograms, boxplots, and scatter plots), descriptive statistics unveil patterns, trends, and potential anomalies within the data. Moreover, this exploratory phase is crucial in identifying data quality issues such as outliers, missing values, or skewed distributions, which might influence subsequent analyses. By distilling the essential structure of the dataset, descriptive statistics not only enhance comprehension but also inform and direct the following stages of inferential statistical modeling and hypothesis testing [12].

Table 1 Statistics Summary

| Variable | Climate Variables | | |
|---|---|---|---|
| | Temperature | Precipitation | Wind Speed |
| Min | 25.360 | 0.0000 | 0.680 |
| Median | 27.150 | 6.540 | 2.020 |
| Mean | 27.146 | 11.425 | 2.164 |
| Max | 28.760 | 133.560 | 5.340 |
| IQR | 0.840 | 11.440 | 0.950 |
| STDV | 0.594 | 14.156 | 0.736 |

Source : (Research Results, 2025)

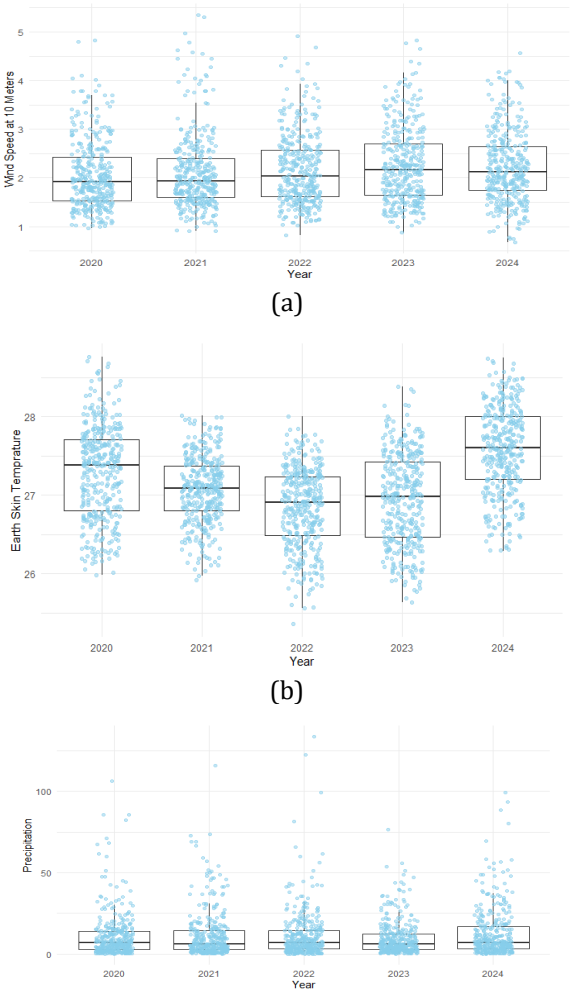
Table 1 indicates that each variable exhibits a unique distribution profile. The surface temperature variable, for instance, has a relatively narrow range (25.36–28.76) and identical mean and median values (27.15), suggesting minimal variation around a central point (standard deviation 0.59). On the other hand, the Precipitation variable spans a substantially wider interval (0.00–133.56). It has a higher standard deviation (14.16), hinting at notable fluctuations that could be tied to seasonal or external influences. The Wind Speed variable falls between these extremes, covering 0.68–5.34 and recording a moderate standard deviation (0.74). Its mean (2.16) is slightly above its median (2.02), reflecting a less concentrated distribution than surface temperature and not as dispersed as Precipitation. These patterns point to different behaviors within the dataset: surface temperature remains relatively stable, Precipitation varies



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

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6665

greatly, and Wind Speed shows an intermediate spread level.




Source : (Research Results, 2025)

Figure 2 (a) Windspeed Boxplot, (b) Earth Skin Temperature Boxplot, and (c) Precipitation Boxplot

Figure 2 illustrates annual distributions (2020–2024) for three essential climate indicators— Precipitation, Surface Temperature, and Wind Speed at 10 Meters. Overall, the median and interquartile ranges imply relatively consistent weather patterns across the years. Nonetheless, occasional outliers in each boxplot point to abnormal or extreme events. Precipitation features several high-value outliers suggestive of intense rainfall episodes, while surface temperature remains steady with minimal fluctuation. Wind speed also demonstrates a generally stable pattern, apart from sporadic spikes. These observations highlight largely stable conditions over the study period, with periodic intervals of more extreme weather.

## A. Stationarity and Granger Causality Test

Ensuring stationarity is crucial for data used in time series analysis, particularly to maintain consistent variance. The Box-Cox method determines the parameter (λ), where a value near 1 indicates variance stationarity. If (λ) deviates significantly from 1, a transformation is performed to achieve this condition. The Augmented Dickey-Fuller (ADF) test is then used to verify mean stationarity.

Table 2 Table of Initial Lambda and Transformation Results

| Variable | Lambda Difference | |
|---|---|---|
| | **λ** Before Transformation | **λ** After Transformation |
| Precipitation | 0.004 | 0.818 |
| Surface Temperature | 1.002 | 1.002 |
| Wind Speed at 10 Meters | 0.213 | 0.896 |

Source : (Research Results, 2025)

Table 2 summarizes the initial and post-transformation (λ) values for three variables: Precipitation, Surface Temperature, and Wind Speed. Initially, (λ) values are distant from 1, suggesting that the variance is not stationary and requires a transformation step. After applying Box-Cox parameters, these (λ) values move closer to 1, signaling that variance stationarity has been established.

Subsequent ADF tests confirm that each variable also meets the criterion for mean stationarity (no differencing needed). As a result, the data satisfy variance and mean stationarity requirements, indicating suitability for modeling and forecasting without additional differencing.

Furthermore, the Granger Causality test was performed to investigate the directional relationships among the analyzed variables Temperature, Precipitation, and Wind Speed. The results reveal that each pair of variables exhibits a bidirectional causal link supported by statistically significant p-values (all $p < 0.05$). Specifically:

1. $Y_2 \rightarrow Y_4$ and $Y_4 \rightarrow Y_2$: Temperature Granger causes Precipitation, and vice versa.
2. $Y_2 \rightarrow Y_6$ and $Y_6 \rightarrow Y_2$: Temperature Granger causes Wind Speed, and vice versa.
3. $Y_4 \rightarrow Y_6$ and $Y_6 \rightarrow Y_4$: Precipitation Granger causes Wind Speed, and vice versa.

These findings indicate that variations in one variable can predict shifts in the others, highlighting the dynamic interplay within the climate system. Such reciprocal influences underscore the importance of jointly examining these variables in

environmental modeling, as temperature, precipitation, or wind speed changes may propagate throughout the system in both directions.

### B. Model Assumption Test on VAR

Various lag orders ranging from 1 to 20 were systematically evaluated to identify the optimal Vector Autoregressive (VAR) model configuration. The selection criterion was based on minimizing the Akaike Information Criterion (AIC), which balances model fit and complexity. Among the evaluated models, the VAR(17) specification yielded the lowest AIC value, indicating its superior performance in capturing the underlying data structure without overfitting. To further validate the adequacy of this model, diagnostic checks were conducted using the Portmanteau test to assess the presence of autocorrelation in the residuals. The VAR(17) model produced a non-significant p-value of 0.6626 ($p > 0.05$), thereby satisfying the white noise assumption. This result confirms that the residuals are independently and identically distributed with no remaining temporal dependencies, reinforcing the model's reliability for inference and forecasting. As a result, no additional lag adjustments or corrective measures were necessary to address residual autocorrelation.

In contrast, alternative lag specifications exhibited significantly lower p-values in the Portmanteau test, suggesting the presence of autocorrelation in their residuals and thereby violating the white noise assumption. Such deficiencies indicate potential model misspecification or underfitting, reducing their suitability for accurate forecasting.

### C. Best Model Selection, Best Model Evaluation and Forecasting

Based on the VAR model selection process, VAR (17) was chosen for forecasting due to its sufficiently low AIC value and its fulfillment of the autocorrelation-free assumption, as demonstrated by the Portmanteau test (p-value = 0.6626). The absence of significant autocorrelation in the residuals makes VAR (17) well-suited for predicting Precipitation, Surface Temprature, and Wind speed.

Table 3 Prediction Evaluation Metrics

| Variable | Evaluation Metrics | | |
|---|---|---|---|
| | MAE | MSE | RMSE |
| Precipitation | 0.177 | 0.043 | 0.208 |
| Surface Temperature | 0.579 | 0.353 | 0.594 |
| Wind Speed at 10 Meters | 0.269 | 0.094 | 0.306 |

Source : (Research Results, 2025)

Table 3 presents the performance metrics MAE, MSE, and RMSE used to evaluate the proposed VAR(17) model across three climate variables: precipitation, surface temperature, and wind speed at 10 meters. The results indicate strong predictive accuracy, with all metrics remaining near zero. Specifically, precipitation exhibits the lowest prediction error (MAE = 0.177; RMSE = 0.208), followed by wind speed (MAE = 0.269; RMSE = 0.306), and surface temperature (MAE = 0.579; RMSE = 0.594). These values demonstrate the model's effectiveness in capturing temporal dependencies across multiple environmental parameters. Following this evaluation, the VAR(17) model was applied to forecast future values using the remaining 20% of the dataset, corresponding to the period from January 1, 2024, through December 31, 2025, spanning 912 days. The consistently low error rates across all variables affirm the model's robustness and reliability for long-term forecasting.

The Bokaa Dam water level prediction study reported higher error margins. Even under optimal conditions, the best-performing dataset (Set-8) yielded an RMSE of 2.7% and MAE of 2.2%, with other sets showing significantly higher RMSE and MAE values. Moreover, some sets produced MAPE values exceeding 50% without climate indices[11]. In contrast, the VAR(17) model achieved lower absolute errors across all climate factors without relying on interpolated or smoothed data, reinforcing its superior generalizability and accuracy.

#### 1. Semi-Supervised Learning

The results of evaluating the optimization of the number of clusters using K-means clustering show that two clusters are the optimal choice, achieving a silhouette score of 0.77. This value indicates optimal data separation, where each cluster exhibits high internal similarity while remaining distinctly apart from others. When the number of clusters increased to three, the silhouette score decreased to 0.73, indicating a decline in separation quality. The trend continued with four clusters, yielding a silhouette score of 0.68, and further declined to 0.65 at five clusters. Based on these observations, two clusters were selected as the most efficient and separable configuration.

These findings are further supported by previous studies emphasizing the strong relationship between high Silhouette Scores and well-defined spherical cluster structures. In numerous studies, a high Silhouette Score is consistently associated with clearly defined spherical clusters. For instance, Abdullah et al.'s research demonstrated the utility of the K-Means


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

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6665

algorithm in identifying optimal cluster formations by achieving a Silhouette Score of 0.608, indicating well-separated clusters of student Instagram accounts [29]. Similarly, Aslantaş et al. reported that their feature set was empirically reduced based on performance measured by the Silhouette Score, with scores exceeding 0.5 interpreted as indicative of reasonable clustering effectiveness [30]. Moreover, empirical evidence supports the premise that spherical clusters yield higher Silhouette Scores due to their inherent geometric characteristics that facilitate a clearer partitioning of space. Kuili et al. elucidated this by explaining that the Silhouette Score quantifies clustering quality based on the concept of distinct separation between clusters, thereby allowing optimal performance in clustering algorithms [31]. Specifically, clustering methods such as K-Means are well-suited for datasets exhibiting spherical characteristics, which enhance the likelihood of achieving superior Silhouette Scores. This synergistic relationship between spherical cluster shapes and elevated Silhouette Scores has been reinforced through various clustering evaluations across different domains [32][33][34].

Furthermore, once the clusters were established, they were interpreted to categorize each day as potentially experiencing an upwelling event. This binary classification laid the groundwork for the subsequent machine learning phase, where a Support Vector Machine (SVM) algorithm was employed to distinguish between upwelling and non-upwelling days. Various SVM kernel functions were evaluated in this phase to determine the most effective classification performance. The results indicated that the Radial Basis Function (RBF) and Polynomial kernels outperformed the Linear kernel, as reflected by their comparatively higher F1 scores.

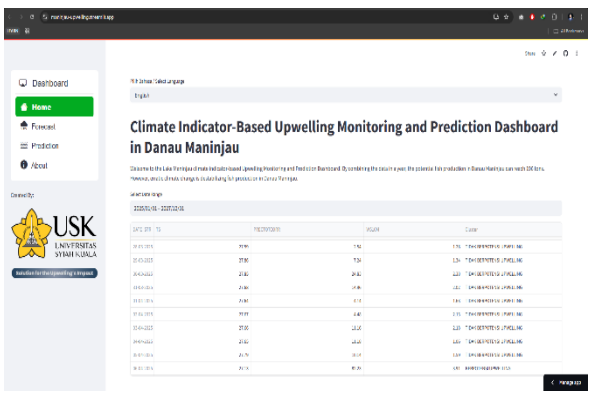
The elevated F1 scores associated with the RBF and Polynomial kernels highlight their superior ability to balance precision and recall, two critical metrics in classification tasks where false positives and false negatives carry significant implications. This suggests that non-linear kernels are more adept at capturing the complex, non-linear relationships inherent in the clustered climatic data. Consequently, these kernels reduce the incidence of misclassification and enhance the model's capability to correctly identify actual upwelling events, thereby increasing the overall robustness and reliability of the predictive framework [12].

D. Upwelling prediction Dashboard

The upwelling prediction dashboard shown in Figure 3 is designed to enhance the awareness and preparedness of floating net cage operators in Maninjau Lake regarding the risk of upwelling events. By selecting a desired time range, the dashboard highlights periods when upwelling may occur, enabling farmers to implement preventive measures and minimize potential losses. The upwelling prediction dashboard is accessible at https://maninjau-upwelling.streamlit.app/.

The importance of developing an effective and user-friendly dashboard is reinforced by findings from previous studies, which underscore the necessity of thorough usability evaluations, particularly in emergency contexts [35]. Heitkemper et al. emphasize that usability testing is often overlooked during dashboard development, yet it is essential for optimizing both design and functionality [35]. In disaster forecasting, where timely responses are critical, dashboards must enable rapid comprehension of complex data through clear visualizations. Usability studies play a crucial role in identifying potential barriers that users may encounter, ensuring that dashboard design maximizes accessibility and minimizes confusion.

Moreover, the responsive design of dashboards has emerged as a decisive factor influencing usability, especially as users increasingly rely on mobile devices to access information during emergencies [36]. Momenipour et al. found that several public health dashboards faced usability challenges on smaller screens, potentially hindering immediate data access in critical situations. Given the increasing prevalence of mobile internet usage, addressing responsive design issues through usability evaluations is vital to ensure that critical weather information is efficiently communicated across various platforms and devices. This approach ultimately enhances the dashboard's functionality and its effectiveness in supporting emergency preparedness and response efforts.




Source : (Research Results, 2025)

Figure 3 Upwelling Prediction Dashboard

Figure 3 shows the dashboard developed using the Streamlit package integrated with Python. Streamlit is recognized as an efficient Python-based solution for rapid model deployment [12]. The deployment process generally involves creating multi-page applications and publishing them through the Streamlight cloud [37], [38] Trained models are stored in .pkl files and delivered via Streamlit applications hosted on GitHub, facilitating broader access [39], [40].

The dashboard presents temperature, precipitation, and wind speed predictions based on existing research regarding upwelling in lakes. Incorporating these scientific insights enables floating net cage farmers to make more informed decisions, contributing to more efficient fish farming management. In addition to serving the needs of Maninjau Lake, the classification model embedded within the dashboard is designed to be adaptable for other lakes. Users can input climate data they have collected, navigate to the prediction section, and manually enter the relevant parameters. By clicking the "Predict" button, the model will classify whether an upwelling event is likely to occur, thus extending the dashboard's applicability beyond its initial deployment site.

## CONCLUSION

This study addressed the critical issue of inadequate prediction capabilities for lake upwelling events at Lake Maninjau, West Sumatra, which pose substantial ecological and socioeconomic threats to local communities reliant on fisheries and tourism. Traditional forecasting methods, typically employing either temporal modeling or event classification independently, were insufficient to accurately predict these events due to the complexity of environmental interactions. A novel hybrid approach integrating Vector Autoregressive (VAR) models with Support Vector Machine (SVM) classifiers was developed and validated to overcome this limitation.

The hybrid VAR(17)-SVM model demonstrated superior predictive performance, achieving notably low forecasting errors with predictive metrics approaching zero and excellent event-classification accuracy, indicated by an F1-score of 0.970 using the SVM with a linear kernel. These findings clearly illustrate the hybrid model's ability to effectively capture temporal dependencies and event-specific characteristics driven by local environmental factors.

Furthermore, the research successfully bridged scientific modeling and community application by developing an interactive, user-friendly dashboard using Python and Streamlit. This dashboard empowers stakeholders to monitor and predict upwelling events in real-time, facilitating informed and proactive resource management decisions. In addition to serving the needs of Lake Maninjau, the classification model embedded within the dashboard is designed to be adaptable for other lakes. Users can input collected climate data, navigate to the prediction section, manually enter the relevant parameters, and activate the "Predict" function. The model will then classify whether an upwelling event is likely to occur, thus extending the dashboard's applicability and supporting broader environmental management initiatives.

Despite the significant improvements presented, there remain opportunities for future research. Subsequent studies could integrate additional environmental parameters, such as dissolved oxygen levels, pH, and nutrient concentrations, which are critical indicators of aquatic ecosystem health. Furthermore, exploring advanced hybrid methods, such as deep learning algorithms or ensemble learning strategies, may further enhance prediction accuracy. Finally, expanding the dashboard to include automated alert systems and mobile-based notifications could improve accessibility and effectiveness, ultimately promoting more resilient and sustainable community-driven lake management strategies.

.

## REFERENCE

[1] A. S. Atmadipoera, A. A. Almatin, R. Zuraida, and Y. Permanawati, "Seasonal upwelling in the northern Arafura sea from multidatasets in 2017," Pertanika J Sci Technol, vol. 28, no. 4, 2020.

[2] A. Nkwasa et al., "Can Turbidity Data from Remote Sensing Explain Modelled Spatial and Temporal Sediment Loading Patterns? An Application in the Lake Tana Basin," Environmental Modeling and Assessment, vol. 29, no. 5, 2024.

[3] Z. A. Haris, A. Irianto, Heldi, R. Dharma, and Yulnafatmawita, "Impact of Natural Disaster on Local Society Income in Maninjau Resort, Agam Regency, Indonesia," Int J Adv Sci Eng Inf Technol, vol. 13, no. 5, 2023.

[4] A. Abrahams, R. W. Schlegel, and A. J. Smit, "Variation and Change of Upwelling Dynamics Detected in the World's Eastern Boundary Upwelling Systems," Front Mar Sci, vol. 8, 2021.

[5] A. Turarbek, M. Bektemesov, A. Ongarbayeva, A. Orazbayeva, A. Koishybekova, and Y. Adetbekov, "Deep Convolutional Neural

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6665

Network for Accurate Prediction of Seismic Events," International Journal of Advanced Computer Science and Applications, vol. 14, no. 10, 2023.

[6] M. S. Kulkarni et al., "Enhancing grid resiliency in distributed energy systems through a comprehensive review and comparative analysis of islanding detection methods," Sci Rep, vol. 14, no. 1, Dec. 2024.

[7] S. A. Valbuena et al., "3D Flow Structures During Upwelling Events in Lakes of Moderate Size," Water Resour Res, vol. 58, no. 3, 2022.

[8] V. Piccialli and M. Sciandrone, "Nonlinear optimization and support vector machines," Ann Oper Res, vol. 314, no. 1, 2022.

[9] P. Wang, X. He, H. Feng, G. Zhang, and C. Rong, "A hybrid model for PM2.5 concentration forecasting based on neighbor structural information, a case in north China," Sustainability (Switzerland), vol. 13, no. 2, 2021.

[10] J. Kairo, "Machine Learning Algorithms for Predictive Maintenance in Manufacturing," 2024. [Online]. Available: www.carijournals.org

[11] Y. O. Ouma et al., "Dam Water Level Prediction Using Vector AutoRegression, Random Forest Regression and MLP-ANN Models Based on Land-Use and Climate Factors," Sustainability (Switzerland), vol. 14, no. 22, 2022.

[12] M. Z. Ulhaq, M. Farid, Z. I. Aziza, T. M. F. Nuzullah, F. Syakir, and N. R. Sasmita, "Forecasting Upwelling Phenomena in Lake Laut Tawar: A Semi-Supervised Learning Approach," Infolitika Journal of Data Science, vol. 2, no. 2, pp. 53–61, Nov. 2024.

[13] S. Oruc, M. A. Hinis, and T. Tugrul, "Evaluating Performances of LSTM, SVM, GPR, and RF for Drought Prediction in Norway: A Wavelet Decomposition Approach on Regional Forecasting," Water (Switzerland), vol. 16, no. 23, Dec. 2024.

[14] NASA, "NASA POWER | Prediction Of Worldwide Energy Resources," 2024. [Online]. Available: https://power.larc.nasa.gov/. [Accessed: 1-January-2025].

[15] P. O. Awodutire, O. R. Ilori, C. Uwandu, and O. A. Akadiri, "Pilot study of new statistical models for prognostic factors in short term survival of oral cancer," Afr Health Sci, vol. 22, no. 2, 2022.

[16] P. O. Awodutire, O. R. Ilori, C. Uwandu, and O. A. Akadiri, "Pilot study of new statistical models for prognostic factors in short term survival of oral cancer," Afr Health Sci, vol. 22, no. 2, 2022.

[17] M. Musyoki, D. Alilah, and D. Angwenyi, "Updated Vector Autoregressive Model Incorporating new Information Using the Bayesian Approach," 2024. [Online]. Available: http://sciencemundi.nethttp://sciencemundi.net

[18] H. Chaudhary, U. Debnath, S. K. J. Pacif, N. U. Molla, G. Mustafa, and S. K. Maurya, "Observational Constraints on the Parameters of Horava-Lifshitz Gravity," Feb. 2024.

[19] A. H. Nasyuha, Zulham, and I. Rusydi, "Implementation of K-means algorithm in data analysis," Telkomnika (Telecommunication Computing Electronics and Control), vol. 20, no. 2, 2022.

[20] J. Mai, "Data-Driven Market Segmentation: K-means Clustering and STP Analysis in Mainland China's Sportswear Industry," International Journal of Global Economics and Management, vol. 4, no. 1, pp. 6–12, Aug. 2024.

[21] H. He, Z. Zhao, W. Luo, and J. Zhang, "Community detection in aviation network based on K-means and complex network," Computer Systems Science and Engineering, vol. 39, no. 2, 2021.

[22] G. Feng, M. Fan, and Y. Chen, "Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining," IEEE Access, vol. 10, 2022.

[23] B. Chong "K-means clustering algorithm: a brief review," Academic Journal of Computing & Information Science, vol. 4, no. 5, 2021.

[24] C. Ioannou, V. Vassiliou, and by Ieee, "Intelligent Systems for the Internet of Things (ISIoT) 2019 workshop, entitled \"Classifying Security Attacks in IoT Networks Using Supervised Learning," 2021.

[25] Suvashisa Dash and Answeta Jaiswal, "Machine learning based forecasting model for rainfall prediction," World Journal of Advanced Research and Reviews, vol. 21, no. 1, 2024.

[26] R. Yoshida, M. Takamori, H. Matsumoto, and K. Miura, "Tropical support vector machines: Evaluations and extension to function spaces," Neural Networks, vol. 157, 2023.

[27] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," 2022.

[28] J. Kairo, "Machine Learning Algorithms for Predictive Maintenance in Manufacturing," 2024. [Online]. Available: www.carijournals.org
[29] A. Abdullah, A. Priadana, M. Muhajir, and S. Nur, "Data mining for determining the best cluster of student Instagram account as new student admission influencer," Telematika, vol. 18, no. 2, p. 255, 2021.
[30] G. Aslantaş, M. Gençgül, M. Rumelli, M. Özsaraç, and G. Bakırlı, "Customer segmentation using k-means clustering algorithm and RFM model," Deu Muhendislik Fakultesi Fen Ve Muhendislik, vol. 25, no. 74, pp. 491–503, 2023.
[31] S. Kuili, K. Dabbour, I. Hasan, A. Herscovich, B. Kantarcı, and M. Chenier, "Adversarial machine-learning-enabled anonymization of OpenWiFi data," WWRT, p. 33–42, 2024.
[32] I. Daniel, L. Akinyemi, and O. Udekwu, "Identifying landslide hotspots using unsupervised clustering: a case study," J. Fut. Artif. Intell. Tech., vol. 1, no. 3, pp. 249–268, 2024.
[33] K. Mondal and J. Klauda, "Physically interpretable performance metrics for clustering,", 2024.
[34] L. Zahrotun, Y. Amanatullah, U. Linarti, and A. Jones, "Strategy for improving and empowering MSMEs through grouping using the AHC method," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 13, no. 1, pp. 130–136, 2024.
[35] E. Heitkemper, S. Hulse, B. Bekemeier, M. Schultz, G. Whitman, and A. Turner, "The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) dashboard for health equity in rural public health: usability evaluation," Jmir Human Factors, vol. 11, p. e51666, 2024.
[36] A. Momenipour, S. Rojas-Murillo, B. Murphy, P. Pennathur, and A. Pennathur, "Usability of state public health department websites for communication during a pandemic: a heuristic evaluation," International Journal of Industrial Ergonomics, vol. 86, p. 103216, 2021.
[37] A. Parker, A. Heflin, and L. C. Jones, "Analyzing University of Virginia Health publications using open data, Python, and Streamlit," J Med Libr Assoc, vol. 109, no. 4, 2021.
[38] E. Schares, "Unsub Extender: A Python-based web application for visualizing Unsub data," Quantitative Science Studies, vol. 3, no. 3, 2022.
[39] J. M. Nápoles-Duarte, A. Biswas, M. I. Parker, J. P. Palomares-Baez, M. A. Chávez-Rojo, and L. M. Rodríguez-Valdez, "Stmol: A component for building interactive molecular visualizations within streamlit web-applications," Front Mol Biosci, vol. 9, 2022.
[40] S. Samanta, M. Pal, R. Mahapatra, K. Das, and R. S. Bhadoria, "A study on semi-directed graphs for social media networks," International Journal of Computational Intelligence Systems, vol. 14, no. 1, 2021.



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