

Estimation of Meteorological Drought in West Kalimantan Using the Random Forest Model

Nurputri, Muhammad Ishak Jumarang*, Andi Ihwan

Geophysics Study Program, Tanjungpura University, Pontianak 78124, Indonesia.

*Corresponding author. Email: ishakjumarang@physics.untan.ac.id

Manuscript received: 25 June 2026; Received in revised form: 7 October 2026; Accepted: 9 October 2026

Abstract

West Kalimantan is highly vulnerable to meteorological drought driven by global climate variability, particularly El Niño–Southern Oscillation (ENSO). This study aims to estimate the Standardized Precipitation Index (SPI) using a Random Forest (RF) model and to analyze the spatio-temporal patterns of meteorological drought across West Kalimantan. Monthly ERA5 reanalysis precipitation data (resolution $\pm 0.25^\circ$) for 1995–2024 were used as the primary dataset; the model was trained on 1995–2018 data and tested on 2019–2024 data. Model performance was evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE). The RF model achieved $R^2 = 0.997$ and $RMSE = 0.052$ during training, and $R^2 = 0.968$ and $RMSE = 0.176$ during testing, demonstrating excellent generalization capability. Spatial analysis of four representative El Niño years (1997, 2009, 2019, 2023) revealed that El Niño consistently triggers dry conditions across the province, with drought intensity scaling with ENSO strength and peak dryness occurring in July–August. The most severe drought was recorded during the Very Strong El Niño of 1997, which caused extreme dryness across most of the province, while Weak and Strong El Niño events of 2019 and 2023 produced less extensive but still significant mid-year drying. These findings demonstrate that the RF model is effective for spatio-temporal drought estimation under varying El Niño intensities and has strong potential to support drought early warning system development in West Kalimantan.

Keywords: Drought; ENSO; Random Forest; SPI; West Kalimantan.

Citation: Nurputri, N., Jumarang, M. I., & Ihwan, A. (2026). Estimation of Meteorological Drought in West Kalimantan Using the Random Forest Model. *Jurnal Geocelebes*, 10(2): 397–412, doi: 10.70561/geocelebes.v10i2.52311

Introduction

Drought is one of the most recurring hydrometeorological disasters, occurring when precipitation falls below the climatological mean over a sustained period, leading to declining river discharge, depleted groundwater reserves, and disrupted food and water security (Mukherjee et al., 2018). Meteorological drought—defined as a prolonged deficit in precipitation relative to the climatological normal—is considered an early indicator of drought onset in Indonesia (Irawan & Komori, 2025). Strong El Niño events, including the Very Strong 1997–1998 event, were associated with severe drought

and massive wildfires in Borneo (Taufik et al., 2017; Ihwan et al., 2019). The impacts of drought include declining river discharge, reduced groundwater availability, and threats to food and water security (Widyastuti, 2020). Although Indonesia is generally recognized as a high-rainfall country, precipitation is unevenly distributed spatially and seasonally. In West Kalimantan, the June–September period is generally drier than other months, while El Niño events can substantially reduce monthly precipitation, contributing to periodic moisture deficits (Carlson et al., 2014).

West Kalimantan, situated on the equatorial belt at 2°05'N–3°05'S and 108°30'–114°10'E, exhibits a humid tropical climate with year-round high rainfall. Nevertheless, the region is particularly susceptible to meteorological drought during El Niño phases, when El Niño-related anomalous circulation over the Maritime Continent reduces moisture transport toward Indonesia (Supari et al., 2018). The regional consequences of dry conditions are also reflected in the historical occurrence of forest and land fires. In Kubu Raya Regency, West Kalimantan, a total of 342 forest and land fire cases were recorded during 2015–2019, with an average of approximately 60 cases per year (Muharrama & Widjonarko, 2023). These recurrent fire events provide quantitative evidence of the vulnerability of the region to fire during periods of dry conditions.

Prolonged precipitation deficits dry out peatlands, heightening the risk of land and forest fires, which in turn generate transboundary haze and serious public-health consequences (Qonita et al., 2019). Despite the acknowledged risk, spatially resolved and temporally continuous drought monitoring in West Kalimantan remains limited.

Meteorological drought—the earliest manifestation of drought along the hydrological cascade—is most commonly quantified by the Standardized Precipitation Index (SPI), developed by McKee et al. (1993). SPI requires only precipitation data, is applicable at multiple time scales, and produces standardized values that allow cross-regional comparison (Yuningsih et al., 2022). However, computing SPI for every grid cell across a large domain demands efficient computational approaches. Machine learning methods, particularly ensemble algorithms, have shown strong capability in modeling complex non-linear hydroclimatic relationships (Roihan et al., 2020).

Effective drought mitigation in West Kalimantan requires accurate spatial data to guide targeted policy responses, yet existing studies leave critical gaps unaddressed. Previous studies have demonstrated the effectiveness of Random Forest (RF) for drought classification and prediction in Indonesia. Ayuningtyas & Prasetyo (2020) reported an accuracy of 99.91% for drought classification in Yogyakarta, while Heo et al. (2023) achieved over 80% accuracy using RF integrated with climate projections in Kalimantan. Silvianugroho et al. (2019) reported an accuracy of 75.07% in Central Java, with SAVI identified as the dominant predictor, while Rezaei et al. (2025) demonstrated further improvement using a hybrid RF–SVR approach. However, these studies have primarily focused on drought classification, risk mapping, or prediction using satellite-derived vegetation indices, climate projections, or other drought-related predictors rather than continuous, grid-scale SPI estimation across West Kalimantan. ENSO classification based on the Oceanic Niño Index (ONI) enables systematic differentiation of drought intensities across El Niño and La Niña phases (BMKG, 2026). However, the application of RF for continuous, grid-scale SPI estimation across the entire West Kalimantan province using ERA5 reanalysis data has received limited attention. Furthermore, the extent to which RF can accurately reproduce drought patterns during neutral ENSO conditions—driven primarily by internal monsoon seasonality—remains insufficiently explored. Addressing these gaps is essential for developing a robust, province-wide drought estimation system.

This study therefore estimates monthly SPI-1 values at every ERA5 grid point in West Kalimantan for the period 1995–2024 using an RF model trained on feature-engineered hydroclimatic variables, and evaluates model performance systematically across six representative

ENSO phases ranging from Very Strong El Niño to Strong La Niña, including a neutral-phase year. SPI-1 was selected because the primary objective of this study is to characterize short-term, monthly precipitation deficits and their spatial variability across West Kalimantan. A one-month accumulation period is particularly suitable for identifying rapid changes in meteorological moisture conditions and their responses to seasonal and ENSO-related precipitation variability. Longer accumulation periods, such as SPI-3, SPI-6, or SPI-12, are more appropriate for assessing medium- to long-term hydrological and agricultural drought conditions, but they can smooth short-term precipitation anomalies. Therefore, SPI-1 was selected to maintain the monthly temporal resolution required for examining drought variability across different ENSO phases. The specific objectives are: (1) to evaluate the accuracy and generalization capability of the RF model in estimating SPI-1 across West Kalimantan, and (2) to characterize the spatio-temporal patterns of meteorological drought as reproduced by the model under different ENSO intensities. The resulting spatio-temporal drought maps are expected to provide a science-based foundation for drought early warning and

hydrometeorological disaster risk management in West Kalimantan.

Materials and Methods

Study Area and Data

The study area covers the entire province of West Kalimantan, Indonesia ($2^{\circ}05'N$ – $3^{\circ}05'S$; $108^{\circ}30'$ – $114^{\circ}10'E$) as shown in Figure 1. Monthly total precipitation data were obtained from the ERA5 reanalysis product (European Centre for Medium-Range Weather Forecasts Reanalysis v5), available via the Climate Data Store (CDS) at <https://cds.climate.copernicus.eu>, with a spatial resolution of $\pm 0.25^{\circ}$ and covering the period January 1995 to December 2024 (360 months). ERA5 was selected because its $0.25^{\circ} \times 0.25^{\circ}$ spatial resolution provides spatially continuous precipitation information suitable for regional drought analysis (Keune et al., 2025). A total of 195 ERA5 grid cells at a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$ were located within the West Kalimantan administrative boundary and included in the analysis. ENSO phase and intensity were classified using the Oceanic Niño Index (ONI) obtained from the NOAA Climate Prediction Center (CPC), defined as the three-month running mean of SST anomalies in the Niño 3.4 region.

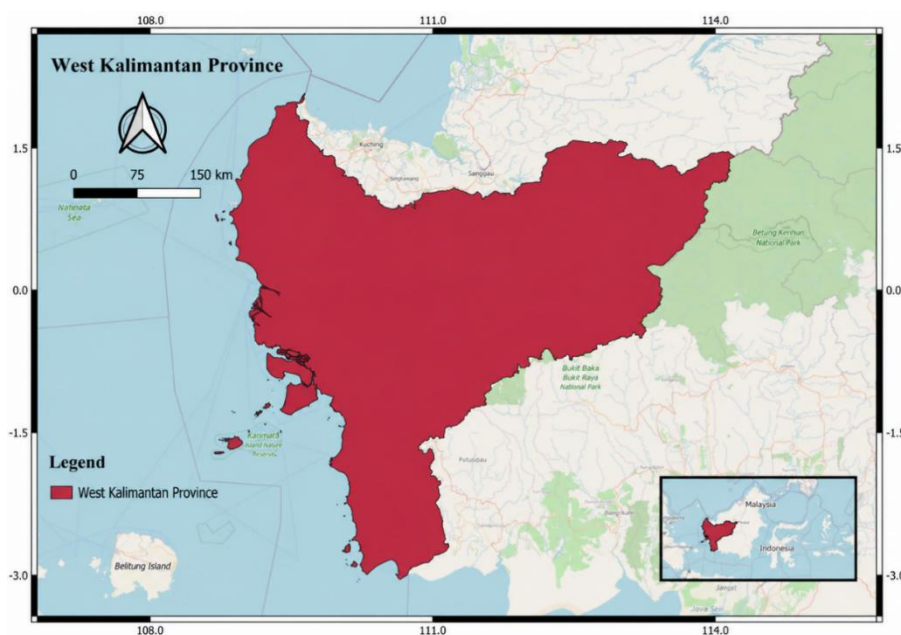


Figure 1. Study location.

SPI Calculation

The one-month Standardized Precipitation Index (SPI-1) was calculated independently for each of the 195 ERA5 grid cells using its corresponding 360-month precipitation time series. Prior to SPI computation, the precipitation time series for each grid cell were checked for missing values. No missing values were identified in the 360-month ERA5 precipitation records; therefore, no imputation or interpolation was required. The Gamma distribution parameters were estimated separately for each grid cell rather than from the pooled dataset, thereby preserving the local precipitation characteristics of each grid location. The SPI computation was implemented in Python on Google Colab (Alfarizi et al., 2023), following the procedure of McKee et al. (1993). SPI is widely used because it requires only precipitation data, is applicable across multiple timescales, and allows spatial comparison between regions with different climatic characteristics (Saidah et al., 2017). For each ERA5 grid cell, non-zero monthly precipitation values were fitted to a two-parameter Gamma distribution (McKee, 1993; Saidah et al., 2017), with the probability density function given in Equation (1):

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, x > 0 \quad (1)$$

where α and β are the shape and scale parameters, respectively. The parameters were estimated separately for each grid cell using the sample mean and the mean of the natural logarithm of non-zero precipitation values, as expressed in Equations (2)-(4) (McKee, 1993; Widyastuti, 2020):

$$A = \ln(\bar{x}) - \frac{1}{n} \sum \ln(x_i) \quad (2)$$

$$\alpha = \frac{1 + \sqrt{1 + \frac{4A}{3}}}{4A} \quad (3)$$

$$\beta = \frac{\bar{x}}{\alpha} \quad (4)$$

where $\bar{x}$ is the mean of non-zero precipitation and n is the number of non-zero observations. Because the Gamma distribution is undefined at $x = 0$, the probability of zero precipitation ($q = m/n$, where m is the count of zero-precipitation months) is incorporated into the cumulative probability as given in Equation (5) (McKee, 1993):

$$H(x) = q + (1 - q)G(x) \quad (5)$$

where $G(x)$ is the Gamma cumulative distribution function. The cumulative probability $H(x)$ was then transformed to the standard normal distribution to obtain SPI-1 using Equation (6) (McKee, 1993) where Φ^{-1} is the inverse of the standard normal cumulative distribution function:

$$\text{SPI} = \Phi^{-1}[H(x)] \quad (6)$$

The resulting SPI values were classified according to McKee et al. (1993), as summarized in Table 1.

Table 1. SPI classification scheme.

SPI Value	Classification
≥ 2.0	Extremely wet
1.50 to 1.99	Very wet
1.00 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.00 to -1.49	Moderately dry
-1.50 to -1.99	Severely dry
≤ -2.0	Extremely dry

Feature Engineering and Random Forest Model

After computing SPI-1 for each grid cell, a feature dataset was constructed for model training. The overall research workflow is summarized in Figure 2. Feature engineering produced the following predictor variables: (1) monthly precipitation (rain), the primary driver of SPI; (2) three-month and six-month rolling mean precipitation (rain_roll3, rain_roll6), capturing accumulated moisture memory; (3) lagged SPI values at one, two, and three months (lag_1, lag_2, lag_3), representing temporal autocorrelation in drought

dynamics. SPI lag variables (lag-1, lag-2, and lag-3) were included to represent the temporal persistence of drought conditions. These variables were derived exclusively from preceding months and were used to capture short-term temporal dependence in SPI. Thus, the lagged predictors provide information available prior to the target month rather than contemporaneous information from the target month.; (4) seasonal cyclic variables using sine and cosine transformations of the calendar month (month_sin, month_cos); and (5) geographic coordinates (latitude, longitude), capturing spatial heterogeneity in climate characteristics across the province.

The dataset was split chronologically: 1995–2018 (approximately 80%) for training and 2019–2024 (approximately 20%) for independent testing. The dataset comprised 221,875 samples, with 176,875 samples (79.72%) used for training and 45,000 samples (20.28%) used for independent testing. The RF model was implemented using Random Forest Regressor from Scikit-learn (Breiman, 2001) with the following hyperparameters: $n_estimators = 600$, $max_features = \text{sqrt}$, $max_samples = 0.8$, $random_state = 42$. These settings were selected to provide a sufficiently large ensemble, introduce randomness among individual trees, and reduce the risk of overfitting while maintaining computational efficiency. Random Forest hyperparameter tuning is the process of finding the best combination of parameters for the Random Forest algorithm to ensure the model delivers optimal predictions while minimizing the risk of overfitting or underfitting (Lujan-Moreno et al., 2018). The training objective is to learn a mapping $f(\cdot)$ from the feature vector X_i to the target SPI value $\hat{y}_i = f(X_i)$, where f aggregates predictions from 600 decision trees through ensemble averaging.

Model Evaluation

Model performance was assessed using two complementary metrics, the coefficient of determination (R^2) and root mean square error (RMSE). R^2 was used to evaluate the proportion of SPI variance explained by the model, while RMSE was used to quantify the magnitude of prediction errors in SPI units. These two metrics were selected to assess both the explanatory performance and prediction error of the Random Forest model. Both R^2 and RMSE were calculated using pooled observations from all grid cells within the study area for the respective training and testing periods, resulting in a single overall value for each metric rather than separate values for individual grid cells. Following Hair et al. (2020), an $R^2 \geq 0.75$ is classified as a very strong relationship. The R^2 and RMSE metrics were calculated using the complete set of observations in the training and testing datasets, combining all grid cells and time steps. Therefore, the reported metrics represent the overall model performance across the study domain rather than performance calculated separately for each grid cell.

The coefficient of determination (R^2) measures the proportion of SPI variance explained by the model as given in Equation (7) (Chicco et al., 2021):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

Root Mean Square Error (RMSE) quantifies mean prediction deviation in SPI units, as given in Equation (8) (Hodson, 2022) where n is the total number of observations:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (8)$$

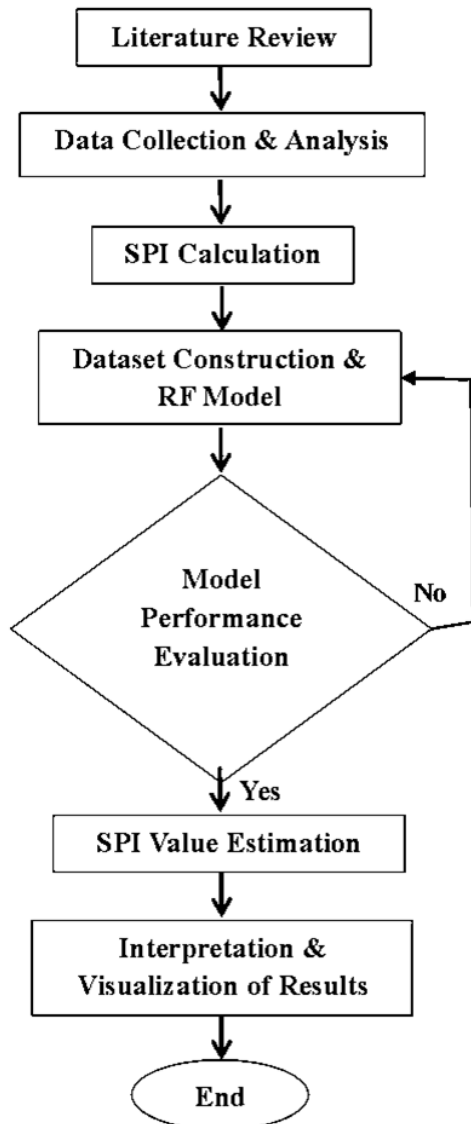


Figure 1. Research flowchart.

Results and Discussion

Results

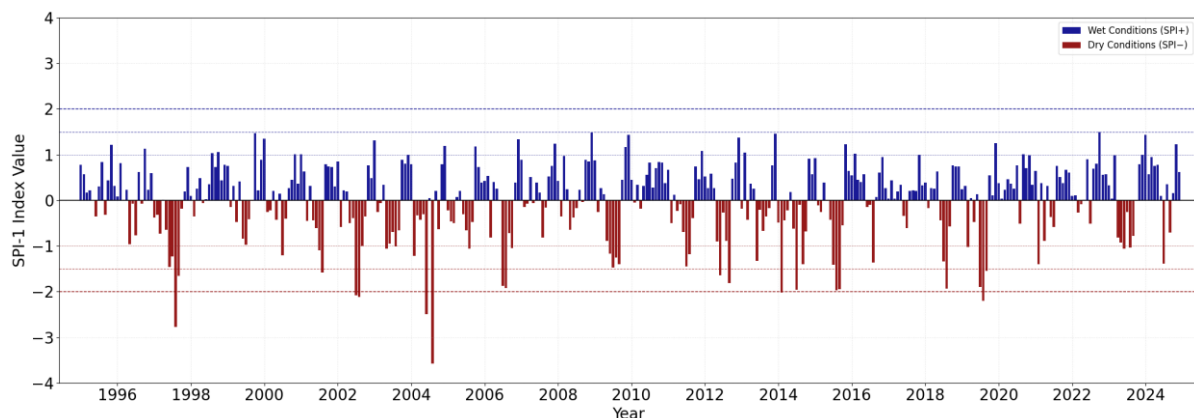


Figure 2. Temporal SPI-1 Time Series for West Kalimantan, 1995–2024.

Analysis of area-averaged SPI-1 across West Kalimantan for 1995–2024 (Figure 3) reveals high interannual variability driven primarily by ENSO. Extremely dry conditions ($\text{SPI} \leq -2.0$) affected approximately 11.6% of grid-month observations in 1997 (SPI ranging from -4.75 to 2.12) and 9.5% in 2015 (SPI ranging from -4.12 to 2.95), coinciding with strong-to-very-strong El Niño events, with the 1997 peak descending below -3.0 in several grid cells. Conversely, extremely wet conditions ($\text{SPI} \geq 2.0$) affected approximately 0.5% of grid-month observations in 2010 (SPI ranging from -1.73 to 2.78) and during 2021–2022, corresponding to La Niña phases.

The RF model trained on 1995–2018 data (Figure 4) achieved $R^2 = 0.997$ and $\text{RMSE} = 0.052$, demonstrating that the model explains 99.7% of SPI-1 variability with minimal average deviation. On the independent test set (2019–2024, Figure 5), performance remained high with $R^2 = 0.968$ and $\text{RMSE} = 0.176$, indicating strong generalization to unseen data. Feature importance analysis (Figure 10) showed that monthly precipitation (rain) contributed 60.1%, followed by rain_roll3 (12.8%), month_cos (7.2%), longitude (4.2%), latitude (3.3%), rain_roll6 (3.1%), and lagged SPI variables collectively (6.9%).

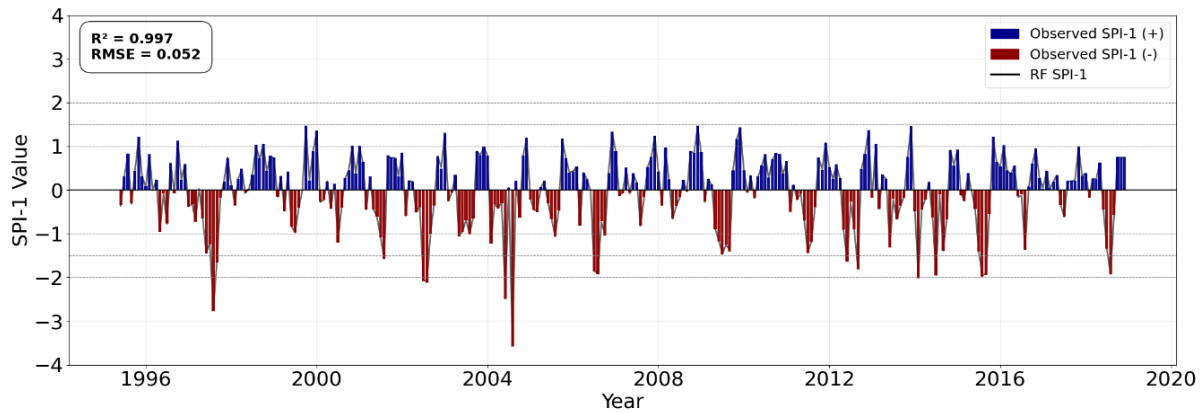


Figure 4. Random Forest Model Estimation on Training Data (1995–2018).

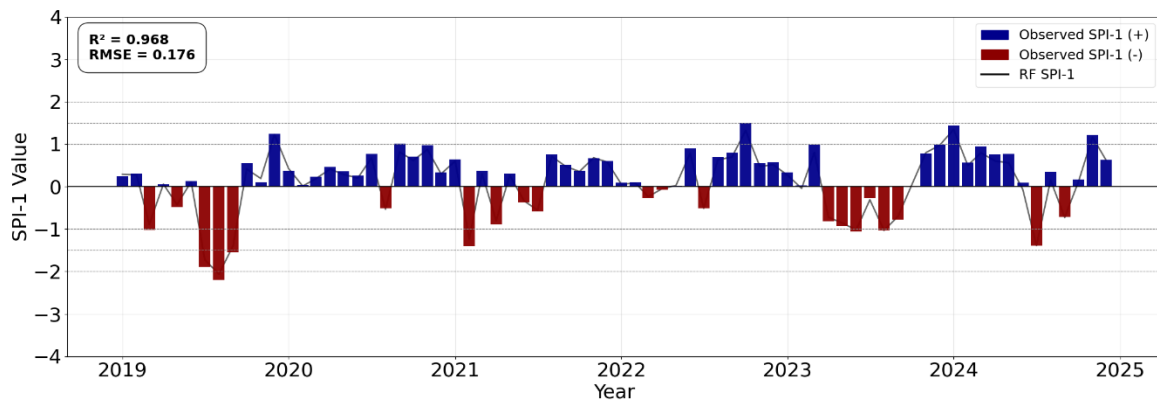


Figure 5. Random Forest Model Estimation on Testing Data (2019–2024).

Model performance evaluated across four representative El Niño years (1997, 2009, 2019, and 2023) is presented in Table 2. The highest R^2 was recorded during Moderate El Niño 2009 ($R^2 = 0.998$), while the lowest was during Strong El Niño 2023 ($R^2 = 0.969$), which belongs to the test set. All R^2 values exceeded 0.96 across all phases. For each year, R^2 and RMSE were computed from all 195 ERA5 grid cells across the 12 months of that year, yielding a sample size of $n = 2,340$ monthly grid-cell observations per year (Table 2).

Table 2. Comparison of R^2 and RMSE values.

Year	ENSO Phase	R^2	RMSE	n
1997	Very strong El Niño	0.997	0.062	2.340
2009	Moderate El Niño	0.998	0.051	2.340
2019	Weak El Niño	0.974	0.189	2.340
2023	Strong El Niño	0.969	0.175	2.340

Spatial distributions of actual SPI-1 (panel A) and RF estimates (panel B) for each representative year are shown in Figures 6–9. During the Very Strong El Niño of 1997 (Figure 6), extremely dry conditions dominated nearly the entire province in July–August, with the most severe deficit concentrated in the southern and southwestern regions. For Moderate El Niño 2009 (Figure 7), a more structured, gradual drought development was observed from May to August, followed by a swift recovery. Weak El Niño 2019 (Figure 8) and Strong El Niño 2023 (Figure 9) both showed clear mid-year drying, though less extensive than 1997, with rapid October recovery.

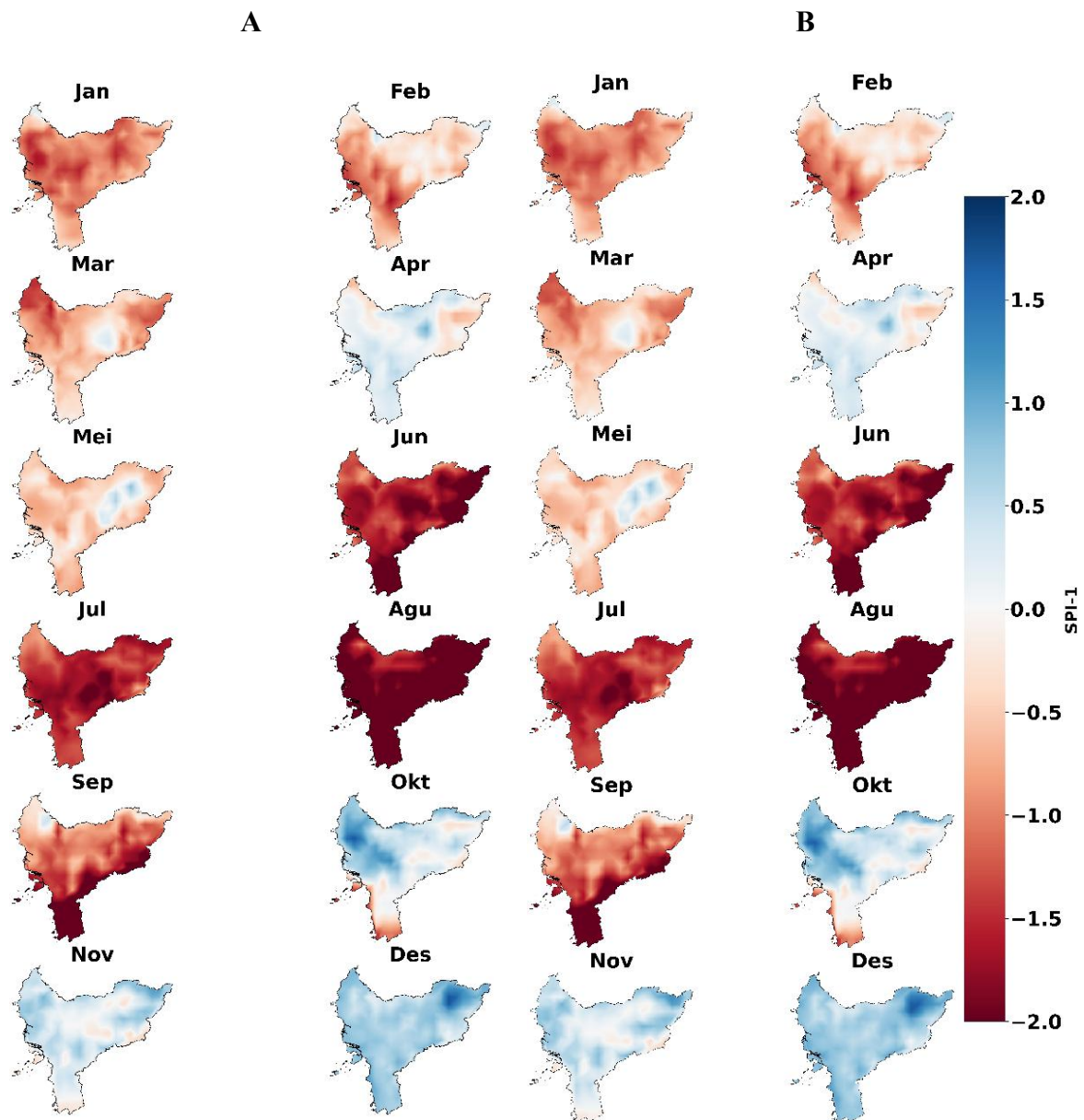


Figure 6. Spatial Distribution of Actual SPI (A) and Random Forest SPI (B) During the Very Strong El Niño Event (Year 1997).

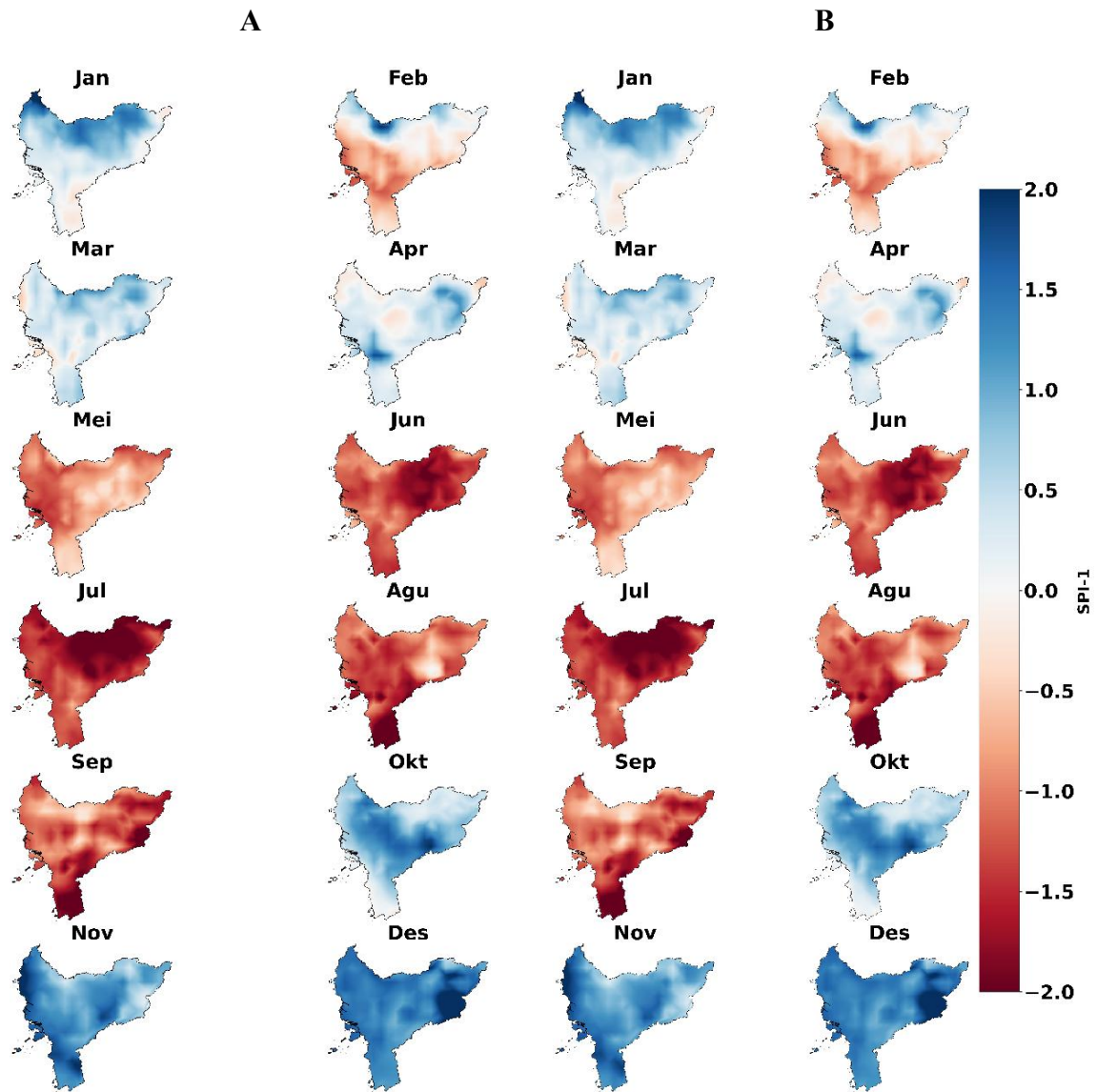


Figure 7. Spatial Distribution of Actual SPI (A) and Random Forest SPI (B) During the Moderate El Niño Event (Year 2009).

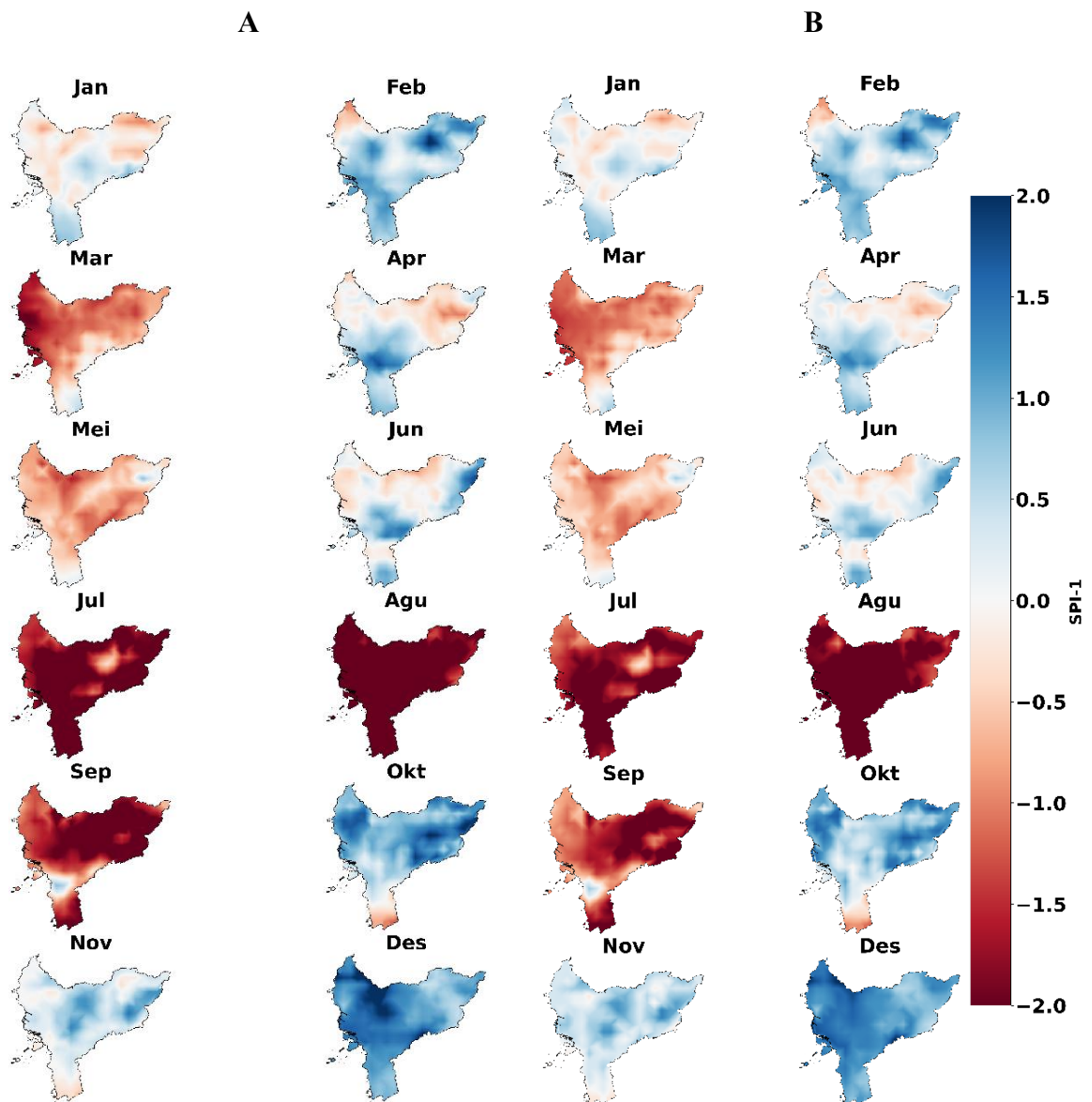


Figure 8. Spatial Distribution of Actual SPI (A) and Random Forest SPI (B) During the Weak El Niño Event (Year 2019).

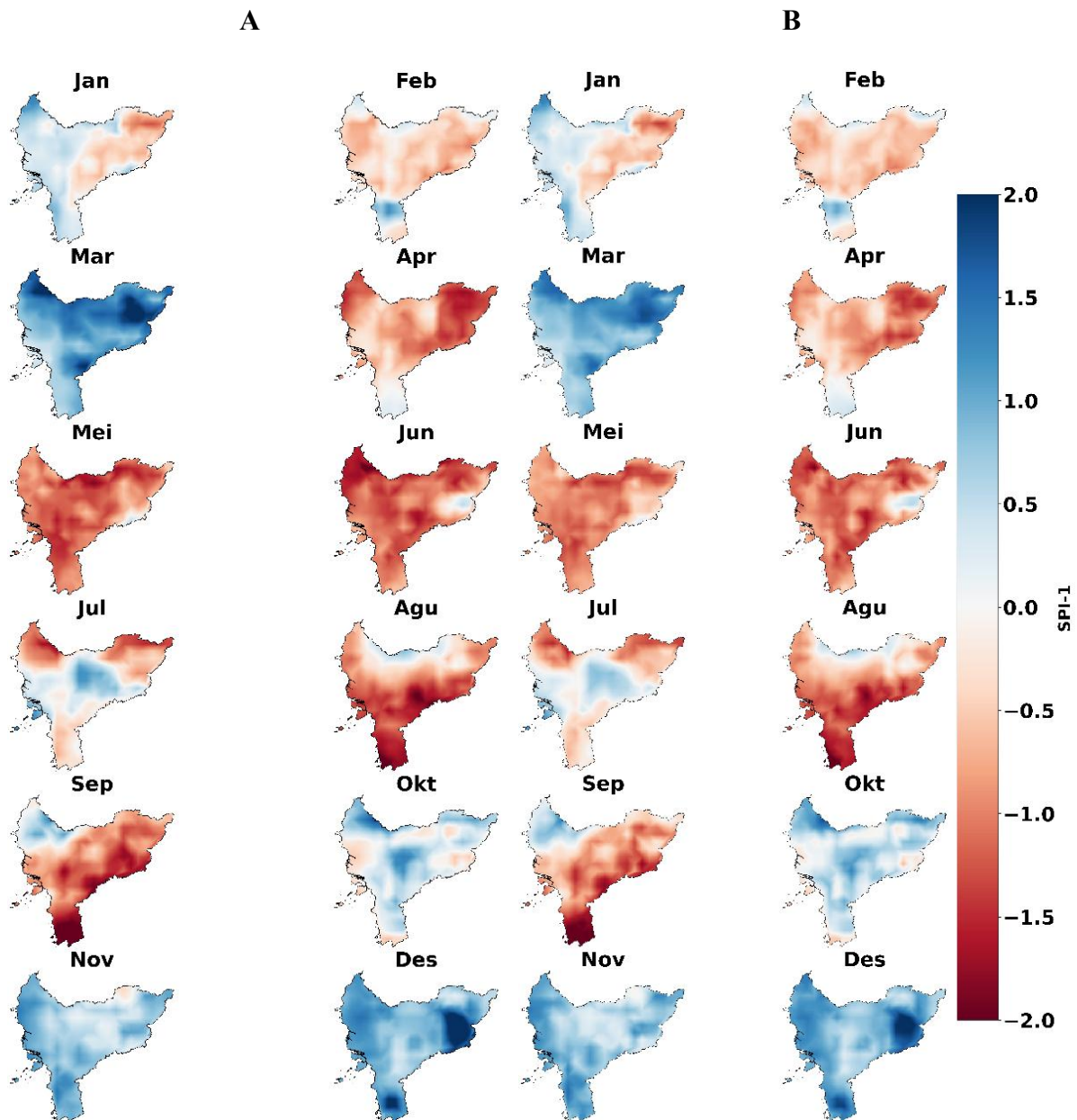


Figure 9. Spatial Distribution of Actual SPI (A) and Random Forest SPI (B) During the Strong El Niño Event (Year 2023).

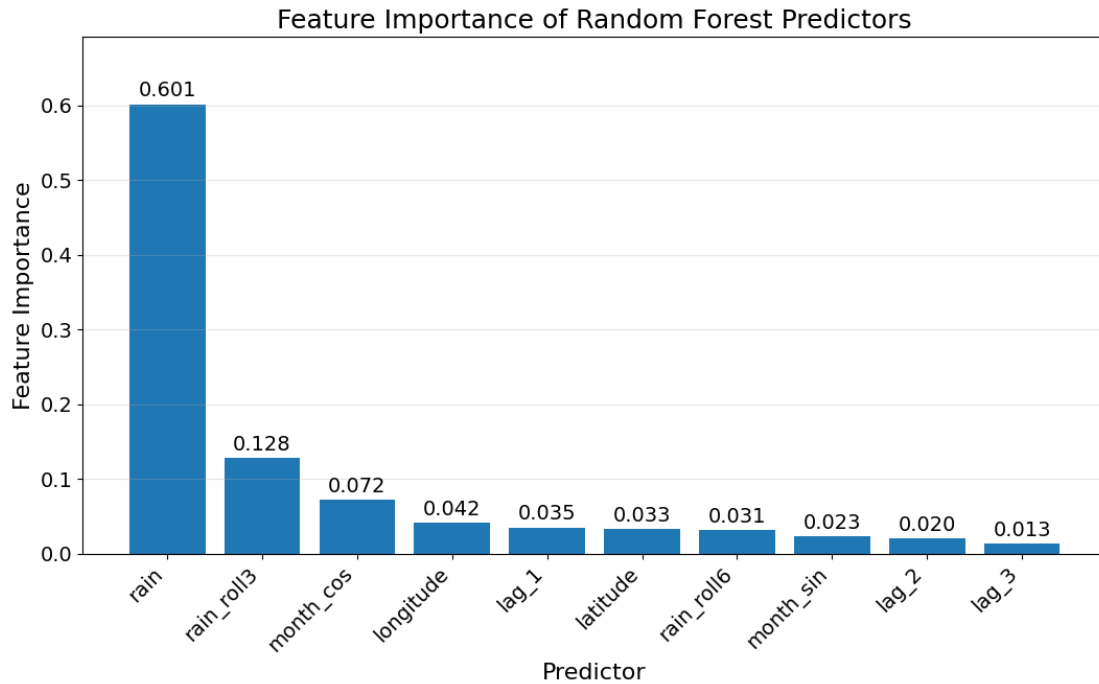


Figure 10. Relative Importance of Predictor Variables in the Random Forest Model for SPI-1 Estimation.

Discussion

The temporal SPI-1 pattern (Figure 3) confirms a strong ENSO–drought coupling in West Kalimantan. The extreme drought of 1997 aligns with one of the strongest El Niño events on record; according to Supari et al. (2018), El Niño suppresses seasonal rainfall over most of Indonesia. This is consistent with Taufik et al. (2017), who linked strong El Niño years to widespread hydrological drought and wildfires in Borneo, as confirmed by Maulida et al. (2025) for the Kalimantan region.

The RF model's strong performance— $R^2 = 0.997$ during training and $R^2 = 0.968$ during testing—demonstrates that ensemble learning effectively captures the non-linear hydroclimatic relationships governing SPI dynamics at the grid scale. The modest R^2 decrease from training to testing ($\Delta = 0.029$) indicates excellent generalization, consistent with the inherent overfitting resistance of RF through bootstrap aggregation (Breiman, 2001). These results compare favorably with previous RF-based drought studies; Ayuningtyas & Prasetyo (2020) achieved 99.91% accuracy in Yogyakarta and Heo et al. (2023) exceeded 80% for multi-hazard risk mapping in

Kalimantan, though both used classification rather than continuous SPI estimation. Because those studies report classification accuracy while the present study reports regression-based R^2 and RMSE for a continuous SPI-1 target, these figures are not directly comparable in a strict numerical sense; the comparison above is intended only to indicate that RF-based approaches generally perform well for drought-related estimation in Indonesia, not to claim superiority on a common metric. This constitutes a limitation of the present study: because no shared regression-based benchmark for RF-based drought estimation currently exists in the Indonesian literature, classification accuracy and continuous SPI regression performance cannot be placed on a single comparative scale, and the high R^2 obtained here should not be interpreted as evidence of superior model skill relative to these prior classification-based studies. The slight underestimation of extreme SPI values in the test set reflects the well-known smoothing effect of ensemble averaging (Rezaei et al., 2025), whereby the mean prediction across many trees regresses toward central values during anomalous conditions.

Feature importance results confirm the fundamental role of precipitation in determining SPI: monthly precipitation alone accounts for 60.1% of the model importance, while the 12.8% contribution of the three-month rolling mean (rain_roll3) reflects accumulated moisture memory. The dominant contribution of precipitation is expected because SPI is directly derived from precipitation and therefore should not be interpreted as a novel finding by itself. Instead, these results confirm that the model captures the expected relationship between precipitation variability and meteorological drought. The contribution of rain_roll3 further indicates that accumulated precipitation information provides additional temporal context for SPI estimation, consistent with the nature of SPI as a transformed precipitation anomaly (McKee, 1993). The combined 6.9% contribution of three lag-SPI variables corroborates the temporal autocorrelation of drought, where past drought conditions influence present ones. Spatial coordinates contribute 7.5% in total, reflecting the heterogeneous climate gradient across the province documented by Putri et al. (2024), who found spatially varying trends in precipitation extremes across Kalimantan.

Spatial comparisons between actual (panel A) and RF-estimated (panel B) SPI-1 maps reveal consistently high agreement across all four El Niño phases. During the 1997 Very Strong El Niño (Figure 6), estimated maps faithfully reproduced the spatial extent and intensity of extreme dryness ($R^2 = 0.997$), particularly the intensification in southern and southwestern areas. This spatial concentration of drought in southern and southwestern West Kalimantan may be related to the seasonal distribution of rainfall and the regional response to El Niño. During El Niño events, the eastward displacement of the Walker circulation suppresses convection and reduces moisture availability over the Indonesian Maritime Continent, particularly during the

dry season. The southern and southwestern parts of West Kalimantan may therefore experience greater rainfall deficits when seasonal moisture supply is reduced, resulting in stronger negative SPI anomalies. Local geographic and topographic characteristics may further modulate rainfall distribution and contribute to the observed spatial differences in drought intensity. For Moderate El Niño 2009 (Figure 7), the model achieved its highest R^2 (0.998), suggesting that moderate-intensity El Niño signals produce more spatially coherent drought patterns that are easier to learn. For the two test-set years—Weak El Niño 2019 (Figure 8, $R^2 = 0.974$) and Strong El Niño 2023 (Figure 9, $R^2 = 0.969$)—the spatial agreement remained high, with only moderate underestimation of peak dry intensities attributable to the smoothing effect of ensemble averaging (Rezaei et al., 2025). The pattern that drought intensity scales with El Niño strength, as evidenced across all four years, corroborates Qonita et al. (2019) finding that ENSO is the primary driver of drought distribution in West Kalimantan.

Conclusion

This study demonstrates that a Random Forest model trained on ERA5 reanalysis data can estimate SPI-1 with high accuracy across West Kalimantan ($R^2 = 0.997$, RMSE = 0.052 training; $R^2 = 0.968$, RMSE = 0.176 testing), maintaining consistent performance across all El Niño intensity phases. Meteorological drought in West Kalimantan is predominantly governed by ENSO, with El Niño triggering widespread dry anomalies that peak in July–September and intensify with increasing ENSO strength. The most severe drought episode occurred during the Very Strong El Niño of 1997, which produced extreme dryness across most of the province, while Moderate El Niño 2009, Weak El Niño 2019, and Strong El Niño 2023 exhibited progressively varying drought extents

consistent with their respective ENSO intensities. The spatial agreement between actual and RF-estimated SPI maps was consistently high ($R^2 > 0.96$) across all four representative El Niño years, confirming the model's effectiveness for spatio-temporal drought characterization. These findings provide a methodological foundation for developing grid-scale drought early warning systems in West Kalimantan and demonstrate the broader potential of ERA5-driven RF models for drought monitoring in data-sparse tropical regions.

The results of this study have practical implications for drought monitoring and early warning in West Kalimantan. The Random Forest model provides continuous SPI-1 estimates by integrating precipitation, temporal SPI persistence, rainfall memory, seasonal, and spatial information. This approach can support the identification of emerging drought conditions and their spatial development before they reach more severe categories. In particular, the observed drought patterns during El Niño years indicate that SPI-based monitoring can be used to identify areas that are more susceptible to rainfall deficits during ENSO events.

The model outputs could be incorporated into an operational drought monitoring framework by updating the predictor variables as new precipitation observations become available and generating spatial SPI estimates at regular intervals. Areas showing increasingly negative SPI values could then be prioritized for closer monitoring and early intervention. Such information may support relevant agencies in planning water-resource management, agricultural preparedness, and drought-response measures. However, operational implementation would require continuous data updates, independent validation under different climatic conditions, and integration with other indicators such as

soil moisture, temperature, and reservoir or river conditions.

Acknowledgements

The authors would like to thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA5 reanalysis data through the Copernicus Climate Data Store (CDS), and NOAA for making the Oceanic Niño Index (ONI) data publicly available. The authors also express sincere gratitude to the reviewers and editors for their constructive comments and suggestions that improved the quality of this manuscript.

Author Contribution

Nurputri: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing – Original Draft, Visualization. Muhammad Ishak Jumarang: Supervision, Conceptualization, Writing – Review & Editing, Validation. Andi Ihwan: Supervision, Writing – Review & Editing, Validation, Resources.

Conflict of Interest

The authors declare no conflict of interest.

References

- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karimah Tauhid*, 2(1), 1–6. <https://ojs.unida.info/karimahtauhid/article/view/7518>
- Ayuningtyas, F., & Prasetyo, S. Y. J. (2020). Pemanfaatan Teknologi Machine Learning Untuk Klasifikasi Wilayah Risiko Kekeringan di Daerah Istimewa Yogyakarta Menggunakan Citra Landsat 8 Operational Land Imager (OLI). *Jurnal Transformatika*, 18(1), 13–24. <https://doi.org/10.26623/transformatika>

- ka.v18i1.2140
- BMKG. (2026). *ENSO (El Niño–Southern Oscillation)*, Badan Meteorologi Klimatologi dan Geofisika. <https://iklim.bmkg.go.id/id/enso/>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Carlson, K. M., Curran, L. M., Ponette-González, A. G., Ratnasari, D., Ruspita, Lisnawati, N., Purwanto, Y., Brauman, K. A. & Raymond, P. A. (2014). Influence of watershed-climate interactions on stream temperature, sediment yield, and metabolism along a land use intensity gradient in Indonesian Borneo. *Journal of Geophysical Research: Biogeosciences*, 119(6), 1110–1128. <https://doi.org/10.1002/2013JG002516>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2020). *Multivariate Data Analysis*. Pearson Education Limited.
- Heo, S., Park, S., & Lee, D. K. (2023). Multi-hazard exposure mapping under climate crisis using random forest algorithm for the Kalimantan Islands, Indonesia. *Scientific Reports*, 13(1), 13472. <https://doi.org/10.1038/s41598-023-40106-8>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Ihwan, A., Pawitan, H., Hidayat, R., Latifah, A. L., & Taufik, M. (2019). Analisis Karakteristik Kekeringan DAS Kapuas Kalimantan Barat Berdasarkan Luaran Global Climate Model. *Positron*, 9(2), 74–80. <https://jurnal.untan.ac.id/index.php/jpositron/article/view/35072/0>
- Irawan, A. N. R., & Komori, D. (2025). Long term variability of agricultural drought's characteristic in Indonesia. *Journal of Agricultural Meteorology*, 81(4), 203–213. <https://doi.org/10.2480/agrmet.D-24-00024>
- Keune, J., Di Giuseppe, F., Barnard, C., da Costa, E. D., & Wetterhall, F. (2025). *ERA5–Drought: Global drought indices based on ECMWF reanalysis*. *Scientific Data*, 12, 616. <https://doi.org/10.1038/s41597-025-04896-y>
- Lujan-Moreno, G. A., Howard, P. R., Rojas, O. G., & Montgomery, D. C. (2018). Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. *Expert Systems with Applications*, 109, 195–205. <https://doi.org/10.1016/j.eswa.2018.05.024>
- Maulida, N. P., Ariska, M., Suhadi, S., & Akhsan, H. (2025). Rainfall Characteristics in Kalimantan Island During ENSO and IOD Phases: Insights from Composite Analysis. *Jurnal Ilmu Fisika | Universitas Andalas*, 17(1), 88–100. <https://doi.org/10.25077/jif.17.1.88-100.2025>
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, 179–183. https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf
- Muharrama, D., & Widjonarko, W. (2023).

- Risiko Bencana Kebakaran Hutan dan Lahan Gambut di Kabupaten Kubu Raya, Provinsi Kalimantan Barat. *Teknik PWK (Perencanaan Wilayah Kota)*, 12(2), 160–170. <https://doi.org/10.14710/tpwk.2023.32816>
- Mukherjee, S., Mishra, A., & Trenberth, K. E. (2018). Climate Change and Drought: a Perspective on Drought Indices. *Current Climate Change Reports*, 4(2), 145–163. <https://doi.org/10.1007/s40641-018-0098-x>
- Putri, J. K., Irfan, M., Akhsan, H., Suhadi, S., & Iskandar, I. (2024). Observed trend of precipitation extreme in Kalimantan. *Global NEST Journal*, 26(9), 06543. <https://doi.org/10.30955/gnj.06543>
- Qonita, I. R., Putra, Y. S., & Kushadiwijayanto, A. A. (2019). Pola Distribusi Kekeringan Menggunakan Metode Standardized Precipitation Index (SPI) di Kalimantan Barat. *Prisma Fisika*, 7(2), 139–142. <https://doi.org/10.26418/pf.v7i2.35574>
- Rezaei, M., Moghaddam, M. A., Piri, J., Azizyan, G., & Shamsipour, A. A. (2025). Drought prediction using advanced hybrid machine learning for arid and semi-arid environments. *KSCE Journal of Civil Engineering*, 29(4), 100025. <https://doi.org/10.1016/j.kscej.2024.100025>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *Jurnal Khatulistiwa Informatika*, 5(1), 490845. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Saidah, H., Budianto, M. B., & Hanifah, L. (2017). Analisa indeks dan sebaran kekeringan menggunakan metode Standardized Precipitation Index (SPI) dan Geographical Information System (GIS) untuk Pulau Lombok. *Jurnal Spektran*, 5(2), 173–179.
- Silvianugroho, S., Prasetyo, S. Y. J., & Hartomo, K. D. (2019). Penentuan Wilayah Resiko Bencana Kekeringan di Jawa Tengah Menggunakan Machine Learning dan Indeks Vegetasi pada Citra Landsat 8 OLI. *Indonesian Journal of Computing and Modeling*, 2(2), 17–24. <https://ejournal.uksw.edu/icm/article/view/2952>
- Supari, S., Tangang, F., Salimun, E., Aldrian, E., Sopaheluwakan, A., & Juneng, L. (2018). ENSO modulation of seasonal rainfall and extremes in Indonesia. *Climate Dynamics*, 51(7), 2559–2580. <https://doi.org/10.1007/s00382-017-4028-8>
- Taufik, M., Torfs, P. J. J. F., Uijlenhoet, R., Jones, P. D., Murdiyarso, D., & Van Lanen, H. A. J. (2017). Amplification of wildfire area burnt by hydrological drought in the humid tropics. *Nature Climate Change*, 7(6), 428–431. <https://doi.org/10.1038/nclimate3280>
- Widyastuti, R., Tambunan, M. P., Taqyuddin, T., & Tambunan, R. P. (2020). Pola Sebaran Kekeringan di Kecamatan Simpenan Menggunakan Metode SPI (Standardized Precipitation Index). *Jurnal Geosaintek*, 6(1), 19–24. <https://journal.its.ac.id/index.php/geosaintek/article/view/7312>
- Yuningsih, Y., Febrianti, D., & Norman, Y. (2022). Analisis Kekeringan Standardized Precipitation Index (SPI) Akibat Pengaruh dari ENSO dan IOD di Wilayah Utara Banten. *Buletin Meteorologi, Klimatologi dan Geofisika*, 2(5), 39–49. https://www.balai2bmkg.id/index.php/buletin_mkg/article/view/42