E-ISSN 2621–086X

Indonesian Journal of Applied Statistics

Vol. 7, No. 1, pp. 14-26, May 2024
Journal homepage: https://jurnal.uns.ac.id/ijas/index
doi: 10.13057/ijas.v7i1.83187

Comparing Monthly Rainfall Prediction in West Sumatra Using SARIMA, ETS, LSTM, and XGBoosting Methods

Fadhil Muhammad Aslam, Fadhli Aslama Afghani*
Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Tangerang, Indonesia

***Corresponding author:** fadhli.aslama.afghani@gmail.com

Submitted: 10-Jan-2024 **Revised:** 22-Jun-2024 **Accepted:** 2-Oct-2024

Abstract. *The West Sumatra Province, serving as the trading center on the island of Sumatra, and boasting various attractive tourist destinations, is not immune to incidents of high precipitation leading to hydro-meteorological disasters such as floods and landslides. Therefore, the accurate prediction of monthly rainfall is crucial to minimize the impacts of high precipitation. This research aims to determine the best method for predicting monthly rainfall using data from 1992 to 2022, which can adequately represent its climatological conditions. The results indicate that the Extreme Gradient Boosting method outperforms the Seasonal Autoregressive Integrated Moving Average (SARIMA), Exponential Smoothing (ETS), and Long Short-Term Memory (LSTM) methods in West Sumatra Province, represented by three weather observation points from the BMKG (Climatology Station of West Sumatra, Maritime Meteorology Station of Teluk Bayur, and Minangkabau Meteorology Station). This method exhibits the lowest error values and the strongest correlation between predicted and actual data. This is evident from the Nash-Sutcliffe Efficiency (NSE) values, which are 0.188214535, 0.613823746, and 0.545734162 (unsatisfactory-satisfactory), as well as the obtained correlation values of 0.472103386, 0.795586268, and 0.743002591 (moderate-strong). However, this method is unable to perfectly capture outlier values. These outliers arise as a result of unusual conditions, such as natural disasters or climate changes, and atmospheric phenomena like El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD), leading to exceptionally high or low precipitation.*

Keywords: *Rainfall; SARIMA; Holt-Winters Exponential Smoothing; LSTM; Extreme Gradient Boosting*

1. INTRODUCTION

The West Sumatra Province, serving as the trading center on the island of Sumatra, and boasting various attractive tourist destinations, is not immune to incidents of high precipitation leading to hydro-meteorological disasters such as floods and landslides [1], [2]. Geographically situated along the western coast of the island of Sumatra and adjacent to the Bukit Barisan Mountains, West Sumatra Province experiences its peak rainfall during the months of September, October, and November [3]. The equatorial rainfall pattern also contributes to the average areas in West Sumatra Province experiencing two peaks in the rainy season, occurring in March and November [4]. On the other hand, the high variability in rainfall contributes to precipitation in Kota Padang, particularly being dominated by local rainfall influenced by the land-sea breeze [5]. High rainfall, as a fundamental aspect, is a leading cause of a series of hydrometeorological disasters in West Sumatra Province, rendering the region highly susceptible to natural disasters and the impacts of climate change. This is evident in the period from 2015 to 2022, where there have been a total of 326 flood events, 383 instances of extreme weather, and 165 landslides in West Sumatra Province [6]. This underscores the urgency for more effective handling and adaptation measures in this region to reduce the risks and impacts of such disasters. Rainfall prediction becomes a critical element in the planning and management of natural resources as well as disaster mitigation. Given the dynamic and current conditions, West Sumatra requires an accurate and effective understanding and prediction of rainfall patterns.

Weather prediction is a complex process that combines various scientific methods to forecast future atmospheric conditions. These methods range from traditional statistical models like ARIMA and SARIMA, which analyze data emphasizing historical patterns, to time series data analysis techniques and additive approaches like exponential smoothing to accommodate seasonal fluctuations. Advances in artificial intelligence have also led to the application of sophisticated methods such as artificial neural networks, LSTM, and machine learning algorithms like XGBoost, capable of efficiently processing large datasets. Accuracy in weather prediction has significant implications in various aspects, making it a crucial element in supporting human activities in daily life.

The SARIMA method is an extension of the ARIMA model that takes into account the seasonal component in time series data. This method is effective for the analysis and prediction of data with clear seasonal patterns and is frequently employed in meteorology for rainfall forecasting [7]. Meanwhile, the Holt-Winters exponential smoothing method, with its additive approach, adjusts for both trend and seasonality in time series. This provides a flexible model for data with significant seasonal fluctuations [8], [9].

Long short-term memory (LSTM) is a type of artificial neural network designed to remember information over long periods, making it ideal for time series-based predictions. The advantage of this method lies in its ability to overcome the vanishing gradient problem that occurs in regular neural networks, making it highly effective in modeling long time sequences [10]. On the other hand, XGBoost is an implementation of the gradient boosting algorithm designed for speed and performance. This algorithm is highly renowned among data scientists for its high accuracy in various prediction and classification tasks [11].

Research on rainfall prediction models in West Sumatra Province has been extensively conducted, one of which is the SARIMA method for forecasting monthly rainfall in Tanah Datar Regency. The results showed consistency during the period from December 2018 to April 2019 [12]. On the other hand, there is a study on monthly rainfall prediction in Padang City during the period 2001-2012 using artificial neural networks with backpropagation training function. The study found that with an increasing number of hidden layers and training data, the prediction results improved, achieving a pattern recognition success rate of 99% [13].

The technology supporting the development of machine learning applies more effective methods in predicting rainfall. This research delves deeper into the latest innovations in weather prediction, particularly the use of long short-term memory (LSTM) and extreme gradient boosting (XGBoost). These methods are expected to provide insights and better results compared to traditional approaches such as seasonal autoregressive integrated moving average (SARIMA) and Holt-Winters exponential smoothing of additive models. By comparing these four methods, this research aims to determine the best approach for predicting rainfall in West Sumatra Province. The outcomes of this study are anticipated to assist policymakers and practitioners in meteorology and disaster management to develop better adaptation and mitigation strategies in West Sumatra Province and other areas with similar.

2. MATERIALS AND METHODS

The researched location is situated in West Sumatra Province, with its astronomical coordinates ranging from 00°54' North Latitude to 30°30' South Latitude, and from 980°36' to 1010°53' East Longitude, as shown in Figure 1:

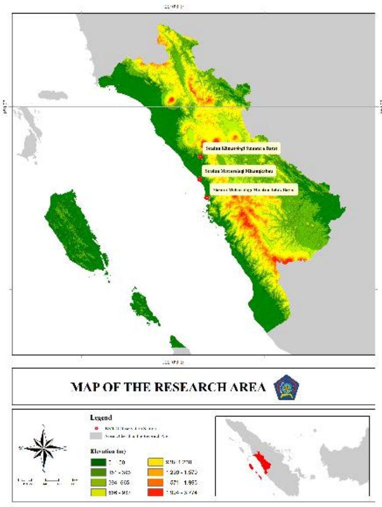

Figure 1. Research location

The research is conducted quantitatively using monthly rainfall data obtained from measurements at BMKG monitoring stations from 1992 to 2022. The data is sourced from the official BMKGSoft website (https://bmkgsoft.database.bmkg.go.id). The data format is CSV with rainfall variable units in the form of millimeters (mm). Missing values in the monthly rainfall data are addressed through the multiple linear regression method, with the equation as follows [14]:

$$y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n \quad (1)$$

with a, b, and y representing the constant, regression coefficient, and dependent variable respectively. Additionally, $b_1, b_2, \ldots, b_n$ and $x, x_2, \ldots, x_n$ denote the regression coefficients and independent variables.

The research study focuses on determining the monthly rainfall prediction method between statistical and machine learning approaches. Data train is 80% of the data length, while data test is 20%. The prediction methods used in this research are as follows:

1. Seasonal Autoregressive Integrated Moving Average (SARIMA)
   SARIMA is a forecasting method that is an extension of the ARIMA model, taking into account seasonal patterns in time series data. This method is easily applied to predict data with seasonal patterns [15]. The mathematical equation for the SARIMA method [16]:

$$\Phi_p B^s \Phi_p(B)(1-B)^d(1-B^s)^D Z_t = \Theta_q(B)\Theta_q(B^s)a_t \quad (2)$$

Where:

p^B : AR non seasonal
$\Phi_p B^s$: AR seasonal
$(1-B)^d$: differencing non seasonal (used to make the time series stationary by removing non-seasonal trends)
$(1-B^s)^D$: differencing seasonal (used to make the time series stationary by removing seasonal trends)

$\Theta_q(B)$: MA non seasonal
$\Theta_q(B^s)$: MA seasonal

Before entering the SARIMA equation, the data needs to be tested for stationary time series using the Augmented Dickey-Fuller (ADF) test. When the p-value results show a value <0.05, then the data can be said to be stationary [17]. Then determine the SARIMA model based on the maximum likelihood estimation (MLE) method with output in the form of autoregressive order 1 (AR1), moving average order 1 (MA1), moving average order 2 (MA2), and seasonal autoregressive order 1 (SAR1) [18], [19]. The results of the MLE will be the basis for forming a SARIMA model to suit the characteristics of the data.

2. Holt-Winters Exponential Smoothing of Additive Models

Holt-Winters exponential smoothing is a combination of grey method and exponential smoothing to perform forecasting used in predicting future events based on past data with trend and seasonal patterns [20]. The equation for the Holt-Winters exponential smoothing additive model with trend and seasonality is as follows [21]:

$$F_{t+m} = S_t + mb_t + I_{t-L+m} \quad (3)$$

with F_t representing the value to be predicted, F_{t+m} being the forecast result for – $(t + m)$, S_t as exponential smoothing at time t, m is the period to be predicted, b_t is the smoothing element at time t, I_t is the seasonal smoothing factor, and L is the length of the season.

3. Long Short-Term Memory (LSTM)

LSTM is a method for modeling time series data resulting from the development of artificial neural networks [22]. The LSTM unit consists of several gates in the form of input, output and forget in each cell which function to regulate the flow of information so that the cell can remember values during changing time intervals [23]. This model has an architecture that can be visualized as follows [24], as shown in Figure 2:

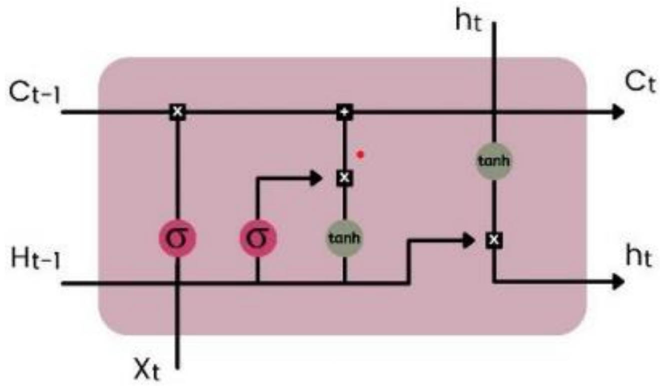

Figure 2. The LSTM architecture

The feedforward calculation for LSTM is as follows [10]:

$$\begin{aligned} f_t &= \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\ c_t &= f_t c_t + i_t \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\ h_t &= o_t * \sigma_h(c_t) \end{aligned} \quad (4)$$

where f_t: forget gate activation vector, σ_g: sigmoid function, W;U: weight matrices to be learned, h_t: output vector of the LSTM unit, b: bias parameter vector to be learned, i_t: input gate activation vector, o_t: output gate activation vector, c_t: cell state vector, x_t: input vector to the LSTM unit, $\sigma_c; \sigma_h$:hyperbolic tangent functions, $*$:Element-wise product (Hadamard product) [25].

4. Extreme Gradient Boosting (XGBoost)
   This model is an ensemble algorithm-based method that is effective in addressing large-scale machine learning cases by leveraging gradient tree boosting, which can prevent overfitting [26]. The objective function of this model is expressed as follows [27]:

$$O = \sum_{j=1}^{T}\left[P_j W_j + \frac{1}{2}\left(Q_j + \eta\right) {W_j}^2\right] + \alpha H \tag{5}$$

Where

p_i	: the first derivative	I	: sample
q_i	: the second derivative α	α	: leaf complexity
P_j	$= \sum_i \epsilon I_j p_i$	H	: the number of leaves
Q_j	$= \epsilon I_j q_i$	η	: penalty parameter
j	: leaf node	${W_j}^2$	: output for each leaf node

The calculation method used in validating the best prediction method is as follows

1. Normalized Root Mean Squared Error (NRMSE)
   The calculation of NRMSE can be used to assess the accuracy of prediction data with the smaller value, better accuracy [28]. The mathematical equation [29]:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{s_i} - P_{g_i}\right)}}{\frac{1}{n}\sum_{i=1}^{n} P_{g_i}} \tag{6}$$

where:
n : number of days or months in the analysis
P_s : predicted data (mm)
P_s : observed rainfall data

2. Nash-Sutcliffe Efficiency (NSE)
   NSE can indicate the level of accuracy in the relationship between observed data and the comparative data with the model. The comparative data is considered good when the NSE value approaches 1 [30], [31]. Mathematical equation [29]:

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Q_{o_i} - Q_{s_i}\right)^2}{\sum_{i=1}^{n}\left(Q_{o_i} - \underline{Q_o}\right)^2}\right] \tag{7}$$

where:
Q_o : Observed rainfall data
Q_s : Predicted rainfall data
$\underline{Q_o}$: Mean observed rainfall data
n : Number of days or months in the analysis
The categories of Nash-Sutcliffe efficiency values, as shown in Table 1.

Table 1. The categories of Nash-Sutcliffe efficiency values [32]

Value	Category
0.75<NSE≤1.00	Very Good
0.65<NSE≤0.75	Good
0.50<NSE≤0.65	Satisfactory
NSE≤0.50	Unsatisfactory

3. BIAS

This method identifies the average response of simulation results when it is underestimated with a negative value and overestimated with a positive value [33]. Bias equation [29]:

$$BIAS = \frac{\sum_{i=1}^{n}\left(P_{s_i} - P_{g_i}\right)}{\sum_{i=1}^{n} P_{g_i}} x100\% \tag{8}$$

Where

n : the number of days or months in the analysis

P_s : the predicted rainfall data (mm)

P_g : the observed rainfall data

4. Pearson Correlation

Pearson correlation is performed when the data exhibits a normal distribution based on the normality test results. This correlation aims to demonstrate the relationship between observed rainfall variables and prediction outcomes. The calculation equation is as follows [34]:

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\left(\sqrt{n\sum X^2 - (\sum X^2)(n\sum Y^2) - (\sum Y^2)}\right)} \tag{9}$$

Where r refers to the correlation coefficient with a range of values between -1 and 1. When the value of r is -1, it indicates a perfect negative correlation, meaning the relationship between variables is inversely proportional. On the other hand, when r is 1, it signifies a perfect positive correlation, indicating a directly proportional relationship between variables. $r = 0$ implies no correlation between variables, as shown in Table 2.

Table 2. Criteria for Pearson correlation coefficient [35]

Value	Category
0-0.19	Very Weak
0.20-0.39	Weak
0.40-0.59	Moderate
0.60-0.79	Strong
0.81-1	Very Strong

5. Mean Absolute Error (MAE)

The smaller the MAE value, the more perfect the prediction data will be to the observation data [36]. The mathematical equation [37]:

$$MAE = \frac{\sum_{n=1}^{N}|\hat{r}_n - r_n|}{N} \tag{10}$$

where $\hat{r}_n$ denotes the ranking of predicted rainfall, r_n represents the actual ranking in the testing dataset or observed rainfall, and N signifies the number of samples. The research methodology used in this study is outlined in Figure 3.

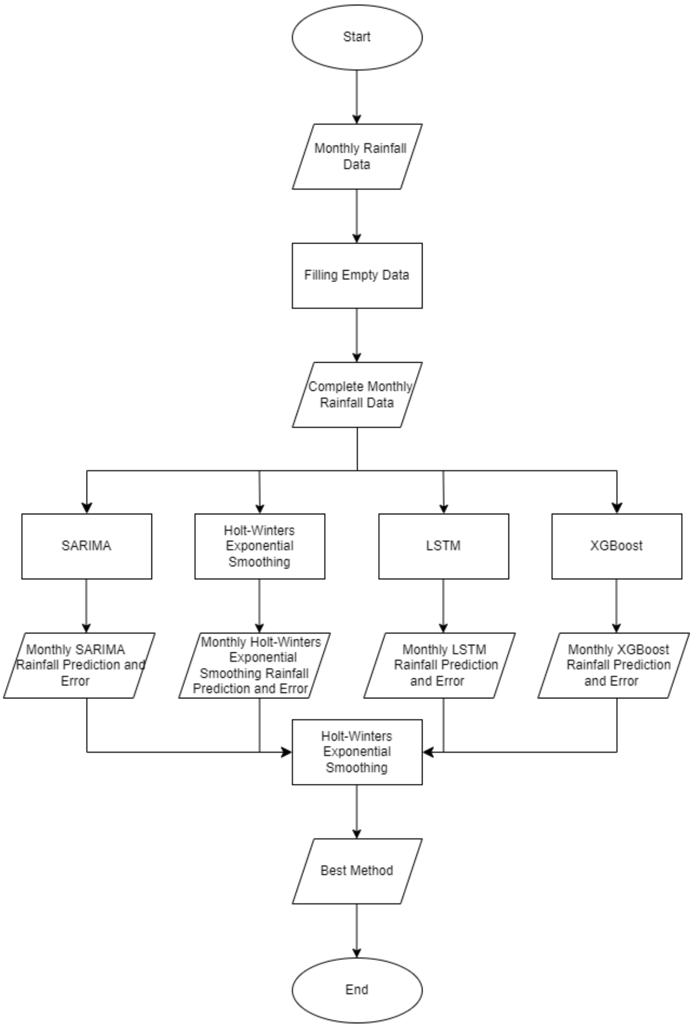

Figure 3. Flowchart

3. RESULTS AND DISCUSSION

3.1 West Sumatra Climatology Station

In predicting rainfall values at the West Sumatra Climatology Station, the XGBoost method stands out as the best among all the reviewed methods. Based on the analysis in Figure 4 and Table 3, XGBoost demonstrates a strong fit with observational data, accurately capturing rainfall fluctuations. In terms of performance, XGBoost has the lowest NRMSE value (0.217), indicating the smallest prediction error among the tested methods. Its positive NSE value (0.188) is one of the highest, indicating good predictive efficiency. Although there is still a tendency for overestimation (BIAS 5.322), it is much lower compared to other methods. The low MAE (113.239) and significant correlation (0.472) confirm the effectiveness of this method.

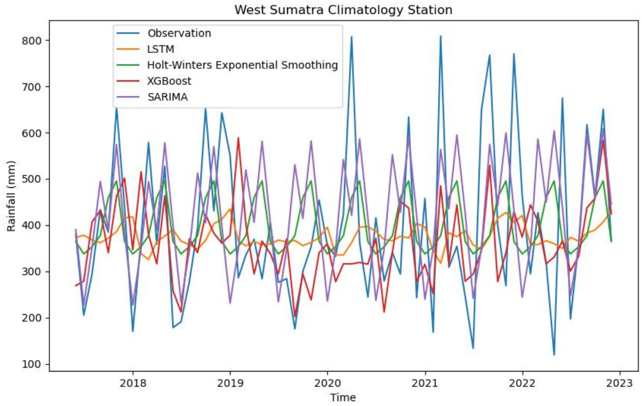

Figure 4. Graph of observed and predicted rainfall at West Sumatra Climatology Station

Table 3. Model performance test

Model	NRMSE	NSE	BIAS	MAE	Correlation
LSTM	0.249	-0.034	15.385	129.308	0.003
Holt-Winters Exponential Smoothing	0.241	-0.005	19.100	131.581	0.169
XGBoost	0.217	0.188	5.322	113.239	0.472
SARIMA	0.254	-0,073	24.582	137.226	0.312

Holt-Winters exponential smoothing, although not as accurate as XGBoost, still demonstrates reasonably good performance. Its predictions follow the observational data better than LSTM but still smooth out the highest peaks. Its NRMSE (0.241) is slightly lower than LSTM, with an almost zero NSE (-0.005), indicating performance not significantly different from average. The BIAS (19.100) and MAE (131.581), slightly higher than LSTM, along with a better correlation (0.169), although still low, suggest a tendency for overestimation.

When examined using the LSTM method, the prediction results from this method follow the trend of observational data but with smaller variations. This indicates a lack of variation in the peaks and valleys of rainfall. LSTM has an NRMSE (0.249) indicating a moderate level of error, a negative NSE (-0.034) indicating its inability to capture data variability as well as the average model, a BIAS (15.385) and MAE (129.308) that are quite high, and a very low correlation (0.003) depicting a weak relationship with observational data.

The SARIMA method, according to the analysis, is recorded as the weakest. Its graph shows sharper and less consistent fluctuations, indicating difficulty in capturing extreme events or sudden variations. SARIMA has one of the highest NRMSE values (0.254), a negative NSE (-0.073), a very large BIAS (24.582), and a high MAE (137.226), all indicating significant prediction errors. The low correlation (0.312) signifies that SARIMA predictions have the weakest relationship with observational data among the methods.

3.2 Maritime Meteorological Station Teluk Bayur

Based on the analysis presented in Figure 5 and Table 4, the XGBoost method exhibits the best performance in predicting rainfall, especially at extreme peaks. The graph indicates the strong ability of XGBoost to adjust to rainfall fluctuations, with the lowest NRMSE (0.144), signifying the smallest prediction errors among all methods. The high NSE (0.613) suggests highly efficient predictive performance, a relatively low BIAS (21.003) indicates a tendency for less overestimation, the lowest MAE (90.137) indicates minimal absolute errors, and a very high correlation (0.795) signifies a very strong relationship with observational data. The SARIMA method, with the SARIMA(2,1,1)(1,1,2) [12] configuration, produces sharp variations in predictions but is less aligned with observational data compared to XGBoost. SARIMA has a relatively high NRMSE (0.218) and a positive NSE (0.121), indicating better efficiency than average but not as strong as XGBoost. The highest BIAS (47.248) in this method indicates a strong tendency for overestimation, a high MAE (137.466), and better correlation than LSTM and Holt-Winters (0.416) but still much lower than XGBoost.

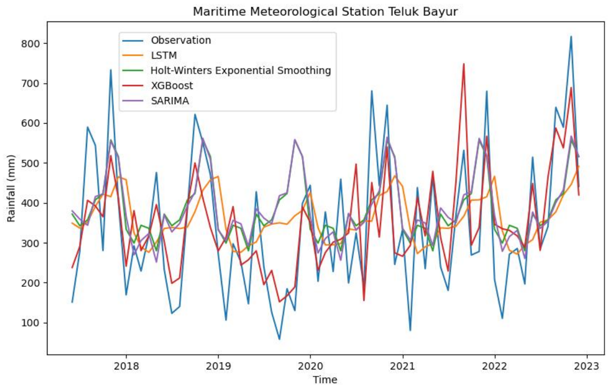

Figure 5. Graph of observed and predicted rainfall at Maritime Meteorological Stasion Teluk Bayur

Table 4. Model performance test

Model	NRMSE	NSE	BIAS	MAE	Correlation
LSTM	0.223	0.074	43.272	142.843	0.302
Holt-Winters Exponential Smoothing	0.223	0.018	38.881	139.150	0.286
XGBoost	0.144	0.613	21.003	90.137	0.795
SARIMA	0.218	0.121	47.248	137.466	0.416

Meanwhile, LSTM and Holt-Winters exponential smoothing show moderate performance. LSTM, with an NRMSE (0.223) reflecting moderate errors, a low NSE (0.074) indicating better but not significant predictive efficiency, a high BIAS (43.272) for overestimation, a large MAE (142.843), and moderate correlation (0.302), shows its ability to follow observational trends but with smaller variations. Holt-Winters, with a similar performance, has an NRMSE comparable to LSTM (0.223), an NSE close to zero (0.018) indicating minimal improvement over the average observation, a slightly lower BIAS (38.881) than LSTM, a slightly lower MAE (139.150), and slightly better correlation (0.286) with observational data.

Other research has also utilized the Markov Chain method to predict ten-day rainfall in eight Rainfall Observation Stations in Padang City. The results indicated that rainfall tends to be high at the Water Plan Semen Observation Station when in a steady state condition, while others experience medium rainfall when reaching a steady state condition. The probability of rainfall in Padang City tends to be moderate [38].

3.3 Minangkabau Meteorological Station

From the analysis conducted in Figure 6 and Table 5, the XGBoost method demonstrates the most accurate performance in predicting rainfall. The graph illustrates XGBoost's strong ability to capture rainfall fluctuations, including extreme peaks and troughs. In the performance table, XGBoost has a very low NRMSE (0.152), indicating the smallest prediction errors among all methods. The very high NSE (0.545) indicates highly efficient predictive performance. The lowest BIAS (9.384) suggests the smallest tendency for overestimation, a very low MAE (87.526) indicates the smallest average absolute error, and a very high correlation (0.743) signifies a very close relationship with observational data.

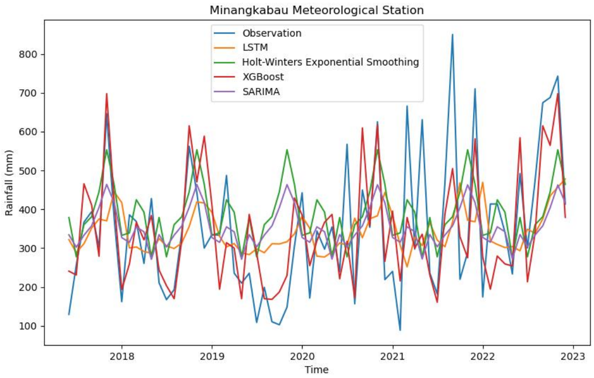

Figure 6. Graph of observed and predicted rainfall at Minangkabau Meteorological Station

Table 5. Method performance test

Model	NRMSE	NSE	BIAS	MAE	Correlation
LSTM	0.229	0.011	28.203	142.310	0.165
Holt-Winters Exponential Smoothing	0.241	-0.005	19.100	131.581	0.169
XGBoost	0.152	0.545	9.384	87.526	0.743
SARIMA	0.217	0.074	27.869	132.354	0.273

SARIMA, with the configuration SARIMA(2,0,1)(1,0,1) [12], exhibits fluctuations in predictions that are less consistent with observational data and struggles to follow extreme peaks. SARIMA has an NRMSE (0.217) larger than XGBoost but smaller than LSTM and Holt-Winters. Its NSE (0.074) suggests that SARIMA performs better than LSTM in capturing predictive efficiency but not as well as XGBoost. A lower BIAS (27.869) compared to LSTM and a higher MAE (132.354) compared to XGBoost, along with a correlation (0.273) indicating a weaker relationship with observational data.

The LSTM model, with an NRMSE of 0.229 indicating moderate errors, exhibits very limited predictive efficiency (NSE 0.011). It has a relatively high BIAS (28.203), large MAE (142.310), and a moderately weak correlation (0.165) with observational data. This indicates its moderate ability to capture patterns in observational data but with smaller variations.

Meanwhile, the Holt-Winters exponential smoothing model shows slightly higher prediction errors compared to LSTM (NRMSE 0.241). NSE approaching zero (0.005) indicates minimal improvement in predictions over the observational average. It has a smaller BIAS (19.100) compared to LSTM, smaller MAE (131.581), and a slightly stronger correlation (0.169) with observational data, but still tends to smooth fluctuations.

However, XGBoost method is unable to perfectly capture outlier values. These outliers arise as a result of unusual conditions, such as natural disasters or climate changes [39] and atmospheric phenomena like El Niño-Southern Oscillation (ENSO) [40] and Indian Ocean Dipole (IOD) [41] leading to exceptionally high or low precipitation.

4. CONCLUSION

Based on the conducted research on statistical and machine learning prediction models, including SARIMA, Holt-Winters exponential smoothing, LSTM, and extreme gradient boosting on monthly rainfall data in West Sumatra Province from 1992 to 2022, representing its climatological conditions. The results indicate that the extreme gradient boosting prediction model has the lowest error values, with the strongest data fitting between actual and predicted values for all tested observation points. This is evident from the correlation values for the three observation points, which are 0.472, 0.795, and 0.743, falling into the moderate-strong category. On the other hand, the NSE values obtained are 0.188, 0.613, and 0.545, categorized as unsatisfactory-satisfactory.

The extreme gradient boosting model's prediction data fails to capture extreme rainfall patterns or outliers. Outliers can occur due to various factors such as unusual conditions like natural disasters or climate change, as well as atmospheric phenomena like El Niño-Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD), resulting in exceptionally high or low rainfall.

ACKNOWLEDGEMENTS

Gratitude is extended to the Badan Meteorologi Klimatologi dan Geofisika for their assistance in providing data for our research.

REFERENCES

[1] A. N. Persada and F. Yustiana, "Analisis Karakteristik dan Klasifikasi Iklim Menurut Schmidt-Ferguson di Provinsi Sumatera Barat," in *FTSP Series: Seminar Nasionald an Diseminasi Tugas Akhir 2023*, FTSP ITENAS, 2023, pp. 909–914.

[2] D. Setiawan, "Analisis Curah Hujan di Indonesia untuk Memetakan Daerah Potensi Banjir dan Tanah Longsor dengan Metode Cluster Fuzzy C-Means dan Singular Value Decompotition (SVD)," *J. EMACS*, vol. 3, no. 3, pp. 115–120, 2021, doi: https://doi.org/10.21512/emacsjournal.v3i3.7428.

[3] E. Aldrian and R. D. Susanto, "Identification Of Three Dominant Rainfall Regions Within Indonesia And Their Relationship To Sea Surface Temperature," *Int. J. Climatol.*, vol. 23, no. 12, pp. 1435–1452, 2003, doi: 10.1002/joc.950.

[4] M. I. Gara, L. Dwiridal, and S. Nugroho, "Analisis Karakteristik Periode Ulang Curah Hujan Dengan Metode Iwai Kadoya Untuk Wilayah Sumatera Barat," *Pillar Phys.*, vol. 12, pp. 47–52, 2019, doi: http://dx.doi.org/10.24036/7515171074.

[5] S. Wahyuni, D. Pujiastuti, L. F. Sani, and A. Rahayu, "Tinjauan Keadaan Meteorologi pada Banjir Bandang Kota Padang Tanggal 24 Juli 2012," *J. Fis. Unand*, vol. 4, no. 4, pp. 405–413, 2015, doi: https://doi.org/10.25077/jfu.4.4.%25p.2015.

[6] Badan Nasional Penganggulangan Bencana (BNPB), "GEOPORTAL DATA BENCANA INDONESIA." Accessed: Nov. 17, 2023. [Online]. Available: https://gis.bnpb.go.id/

[7] M. Geurts, G. E. P. Box, and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," *J. Mark. Res.*, vol. 14, no. 2, 1977, doi: 10.2307/3150485.

[8] C. C. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages," *Int. J. Forecast.*, vol. 20, no. 1, pp. 5–10, 2004, doi: https://doi.org/10.1016/j.ijforecast.2003.09.015.

[9] P. R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," *Manag. Sci*, vol. 6, no. 3, 1960, doi: 10.1287/mnsc.6.3.324.

[10] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Dec. 1997, doi: 10.1162/neco.1997.9.8.1735.

[11] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[12] F. Aulia, H. Yozza, and D. Devianto, "Peramalan Curah Hujan Bulanan Kabupaten Tanah Datar Dengan Model Seasonal Autoregressive Integrated Moving Average (SARIMA)," *J. Mat. UNAND*, vol. 8, no. 2, pp. 37–44, 2019, doi: https://doi.org/10.25077/jmu.8.2.37-

44.2019.
[13] C. Oktaviani and Afdal, “Prediksi Curah Hujan Bulanan Menggunakan Jaringan Syaraf Tiruan Dengan Beberapa Fungsi Pelatihan Backpropagation (Studi Kasus: Stasiun Meteorologi Tabing Padang, Tahun 2001-2012),” *J. Fis. Unand*, vol. 2, no. 5, pp. 228–237, 2013, doi: https://doi.org/10.25077/jfu.2.4.%25p.2013.
[14] W. A. L. Sinaga, S. Sumarno, and I. P. Sari, “Penerapan Metode Regresi Linier Berganda Untuk Estimasi Jumlah Penduduk Pada Kecamatan Gunung Malela,” *J. Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 55–64, 2022.
[15] A. Assidiq, P. Hendikawati, and N. Dwidayati, “Perbandingan Metode Weighted Fuzzy Time Series, Seasonal ARIMA, dan Holt-Winter’s Exponential Smoothing untuk Meramalkan Data Musiman,” *UNNES J. Math.*, vol. 6, no. 2, pp. 129–142, 2017, doi: 10.15294/UJM.V6I2.12288.
[16] R. Fahrudin and I. D. Sumitra, “Peramalan Inflasi Menggunakan Metode Sarima dan Single Exponential Smoothing (Studi Kasus: Kota Bandung),” *Maj. Ilm. UNIKOM*, vol. 17, no. 2, pp. 111–120, 2020, doi: https://doi.org/10.34010/miu.v17i2.3180.
[17] M. Tiyani and A. I. Achmad, “Uji Stasioneritas Dickey Fuller Generalized Least Square pada Data Inflasi Indonesia dari Bulan Januari 2006 Hingga Bulan Mei 2020,” in *Prosiding Statistika,* SPeSIA, 2020, pp. 27–33. doi: http://dx.doi.org/10.29313/.v6i2.22587.
[18] N. Nirwana, M. Hadijati, and N. Fitriyani, “Estimasi Parameter Model Moving Average Orde 1 Menggunakan Metode Momen dan Maximum Likelihood,” *Eig. Math. J.*, vol. 1, no. 1, p. 17, Jun. 2018, doi: https://doi.org/10.29303/emj.v1i1.8.
[19] M. Jannah and I. N. F. H, “Penaksiran Parameter Model Autoregressive Orde (1) Dengan Menggunakan Metode Likelihood Maksimum,” *Math. Appl. J.*, vol. 1, no. 2, pp. 38–48, 2019, doi: https://doi.org/10.15548/map.v1i2.1180.
[20] H. W. Tresnani, A. Sihabuddin, and K. Mustofa, “Optimasi Parameter Pada Metode Peramalan Grey Holt - Winter Exponential Smoothing Dengan Golden Section,” *Berk. MIPA*, vol. 25, no. 3, pp. 312–325, 2018.
[21] Nurhamidah, Nusyirwan, and A. Faisol, “Forecasting Seasonal Time Series Data using The Holt-Winters Exponential Smoothing Method of Additive Models,” *J. Mat. Integr.*, vol. 16, no. 2, pp. 151–157, 2020.
[22] L. Wiranda and M. Sadikin, “Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. Metiska Farma,” *J. Nas. Pendidik. Tek. Inform.*, vol. 8, no. 3, pp. 184–196, 2019, doi: https://dx.doi.org/10.23887/janapati.v8i3.19139.
[23] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning : A systematic literature review: 2005–2019,” *Appl. Soft Comput.*, vol. 90, p. 106181, 2020, doi: https://doi.org/10.1016/j.asoc.2020.10618.
[24] D. D. Pramesti, D. C. R. Novitasari, F. Setiawan, and H. Khaulasari, “Long-Short Term Memory (LSTM) For Predicting Velocity And Direction Sea Surface Current on Bali Strait,” *J. Ilmu Mat. dan Terap.*, vol. 16, no. 2, pp. 451–462, 2022, doi: https://doi.org/10.30598/barekengvol16iss2pp451–462.
[25] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017, doi: 10.1109/TNNLS.2016.2582924.
[26] S. E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, “Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit,” *J. Math. Theory Appl.*, vol. 4, no. 1, pp. 21–26, 2022.
[27] P. S. Rizky, R. H. Hirzi, and U. Hidayaturrohman, “Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang,” *J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 228–236, 2022, doi: https://doi.org/10.36456/jstat.vol15.no2.a5548.
[28] E. D. Amelia, S. Wahyuni, and D. Harisuseno, “Evaluasi Kesesuaian Data Satelit sebagai

Alternatif Ketersediaan Data Evaporasi di Waduk Wonorejo," *J. Tek. Pengair. J. Water Resour. Eng.*, vol. 12, no. 2 SE-Articles, pp. 127–138, Dec. 2021, doi: 10.21776/ub.pengairan.2021.012.02.05.

[29] F. Su, Y. Hong, and D. P. Lettenmaier, "Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin," *J. Hydrometeorol.*, vol. 9, no. 4, pp. 622–640, 2007.

[30] Indarto, *Hidrologi Dasar Teori dan Contoh Aplikasi Model Hidrologi*. Jakarta: Bumi Aksara, 2010.

[31] D. P. Jarwanti, E. Suhartanto, and J. S. Fidari, "Validasi Data Curah Hujan Satelit TRMM (Tropical Rainfall Measuring Mission) dengan Data Pos Penakar Hujan di DAS Grindulu, Kabupaten Pacitan, Jawa Timur," *J. Teknol. dan Rekayasa Sumber Daya Air*, vol. 1, no. 2, pp. 772–785, 2021.

[32] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model Evaluation Guidelines For Systematic Quantification Of Accuracy In Watershed Simulations," *Am. Soc. Agric. Biol. Eng.*, vol. 50, no. 3, p. 885−900, 2007.

[33] D. N. Moriasi, M. W. Gitau, N. Pai, and P. Daggupati, "Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria," *Am. Soc. Agric. Biol. Eng.*, vol. 58, no. 6, pp. 1763–1785, 2015, doi: 10.13031/trans.58.10715.

[34] N. Tallamma, N. Ihsan, and A. J. Patandean, "Analisis Pengaruh Madden Julian Oscillation (MJO) Terhadap Curah Hujan Di Kota Makassar," *J. Sains dan Pendidik. Fis.*, vol. 12, no. 3, pp. 324–329, 2016.

[35] D. A. R. Oktaverina, E. Suhartanto, and S. Wayhuni, "Validasi Data Curah Hujan Satelit TRMM (Tropical Rainfall Measuring Mission) dengan Pos Stasiun Hujan pada Sub DAS Keduang Kabupaten Wonogiri, Provinsi Jawa Tengah," *J. Teknol. dan Rekayasa Sumber Daya Air*, vol. 2, no. 1, pp. 265–276, 2022.

[36] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient of Cetermination R-squared is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *PeerJ. Comput. Sci.*, vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.

[37] W. Wang and Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," in *ICMEMSCE*, IOP Conf. Series: Materials Science and Engineering, 2018. doi: 10.1088/1757-899X/324/1/012049.

[38] U. F. Arshinta and D. Ahmad, "Analisis Curah Hujan di Kota Padang dengan Menggunakan Rantai Markov," *UNPjoMath*, vol. 2, no. 4, pp. 45–50, 2019, doi: http://dx.doi.org/10.24036/unpjomath.v4i4.7928.

[39] Centers for Disease Control and Prevention, "Precipitation Extremes: Heavy Rainfall, Flooding, and Droughts." Accessed: Jan. 05, 2024. [Online]. Available: https://www.cdc.gov/climateandhealth/effects/precipitation_extremes.htm

[40] Sitompul, Zulfahmi, and E. Nurjani, "Pengaruh El Nino Southern Oscillation (ENSO) Terhadap Curah Hujan Musiman Dan Tahunan Di Indonesia," *J. Bumi Indones.*, vol. 2, no. 1, 2013.

[41] A. Kurniadi, E. Weller, S.-K. Min, and M.-G. Seong, "Independent ENSO and IOD Impacts on Rainfall Extremes Over Indonesia," *Int. J. Climatol.*, vol. 41, pp. 1–17, Feb. 2021, doi: 10.1002/joc.7040.