Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:97
https://doi.org/10.28919/cmbn/8820
ISSN: 2052-2541

MULTI-OUTPUT MACHINE LEARNING REGRESSION FOR CLIMATE PREDICTION: A COMPARATIVE STUDY OF PRECIPITATION AND TEMPERATURE FORECASTING IN JAKARTA AND EAST KALIMANTAN, INDONESIA

KARLI EKA SETIAWAN*, HAFIZH ASH SHIDDIQI, PANDU WICAKSONO

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Abstract: The Indonesian government is currently transferring the national capital from Jakarta, on Java Island, to East Kalimantan, on Kalimantan Island. Forecasting climate conditions in both Jakarta and East Kalimantan is crucial for urban planning, disaster risk management, environmental conservation, and economic stability. An accurate climate forecast can guide sustainable development, improve disaster preparedness, and support agriculture, fisheries, and public health. This research focuses on the development of machine learning models for making an accurate climate forecast by proposing some models such as linear regression, polynomial regression, decision tree regressors, K-nearest neighbors regressors, elastic networks, Lasso, and support vector regressors. The predictive models built in this research are multi-output regression for forecasting four climate variables in the next one month as the target or output by using the previous 24 months of data on four climate variables as a feature or input. This research utilized public climate conditions in Jakarta and East Kalimantan obtained from the World Bank Climate Change Knowledge Portal (CCKP). The results show that the linear regression model was the best model for both forecasting scenarios in Jakarta and East Kalimantan.

Keywords: multi-output regression; Indonesia climate forecasting; IKN climate; machine learning for climate forecasting; linear regression; polynomial regression; decision tree regressor; K-nearest neighbors regressor; elastic-net; Lasso; support vector regressor; sustainable urban planning; disaster risk management.

2020 AMS Subject Classification: 68T01, 68T10, 97P80

---

*Corresponding author
E-mail address: karli.setiawan@binus.ac.id

Received May 21, 2011

1. INTRODUCTION

The Indonesian government is currently transferring the national capital from Jakarta on Java Island to East Kalimantan on Kalimantan Island [1]. The reason for this shift is that Jakarta, located on the island of Java, is grappling with numerous sustainability issues, including being the world's fastest sinking city and the most densely populated city [2]. At present, the Indonesian government has extensively constructed infrastructure in the New National Capital (IKN) city. However, there is a pressing need to monitor and predict the climate conditions to facilitate rapid infrastructure development, address agricultural concerns, manage water resources, and prepare for disasters. Another problem is that climate change is causing shifts in weather patterns, making it increasingly difficult to predict using conventional methods [3].

Conducting research in forecasting future atmospheric data can bring benefits for human life and its environment, such as integrated decision-making with other fields like agriculture, water management, and disaster preparedness, and holistic understanding where researchers and policymakers can gain a more holistic understanding of climatic conditions and their potentials [4]. Climate conditions and climate change have a significant impact on the attractiveness of tourism, and weather forecasting can enhance this attractiveness by providing tourists with information on convenience and safety [5]. Forecasting climatology data can also benefit infrastructure construction by safeguarding workforce activities and resources from the dangers of climate change. Weather forecasting aids energy management by providing precise information to identify an optimal and efficient power source, such as renewable energy sources like wind and solar power plants [6].

This research proposes simultaneous forecasting by predicting all variables simultaneously using multi-output regression machine learning (ML) approaches, where we predict all climate variables at one month in the future using all previous 24 months climate variables using ML models such as linear regression, polynomial regression with 2- and 3-degree, decision-tree regressor, K-nearest neighbor (KNN) regressor, Lasso, elastic-net regression, and support vector regression. This research observed two regions in Indonesia, Jakarta and East Kalimantan using public dataset provided by World Bank Data due to East Kalimantan become the new capital city of Indonesia and Jakarta was the capital city of Indonesia.

2. RELATED WORKS

Researchers conducted a review on the development of weather prediction and climate analysis [7]. They concluded that the machine learning (ML) approach will be a key feature in the future development of weather forecasting. According to their research, wind, precipitation, temperature, pressure, and radiation are the most commonly studied in meteorological research, and deep learning, random forest, artificial neural network (ANN), support vector machines (SVM), and Extreme Gradient Boosting (XGBoost) are the most popular models to use for forecasting purposes. This research was done by Bochenek et al. inspired this research to use some climate condition parameter-related temperature and precipitation and explore the capacity of various ML techniques. Another review of ML approaches for numerical weather and climate modeling by Burgh-Day et al. stated that ML approaches for weather and climate modeling are relatively cheap in computational terms compared to current physics-based modeling systems [8]. Despite its use for nearly 30 years, the application of ML in weather and climate research has rapidly accelerated in recent years.

Ahmed et al. conducted a similar study, proposing a multi-model ensemble (MME) that uses ML to predict precipitation and air temperature in Pakistan [9]. They combined various ML models, such as ANN, KNN, SVM, and relevance vector machine (RVM), to build an MME architecture model, and the results showed that the RVM-based MME was the best model. Another researcher conducted a ML experiment for predicting precipitation in the Yalong River Basin in China [10]. They compared 24 ML models to solve their problems, and the optimized extra-trees regressor was the best model. They also concluded that the selection of forecast predictors plays a significant role. Based on the literature review done by Cifuentes et al., in ML research for forecasting air temperature, it can be found that ANN and SVM are the most popular ML models utilized in this topic research, while they also found that the deep learning approach has shown great promise due to its great performance [11]. They also stated that air temperature is a key factor in climate impact studies.

3. RESEARCH METHODOLOGY

This research focused on the comparison study of various popular ML algorithms for forecasting some environmental variables in two cities in Indonesia, Jakarta and East Kalimantan, such as precipitation (pr), average mean surface air temperature (tas), average maximum surface air temperature (tasmax), and average minimum air temperature (tasmin). Intuitively, this research chose to utilize data from the previous two years or 24 months to predict the upcoming month, as shown in Figure 1. Therefore, Figure 2 illustrates the need to implement a sliding window in this research as our previous work [12] [13].

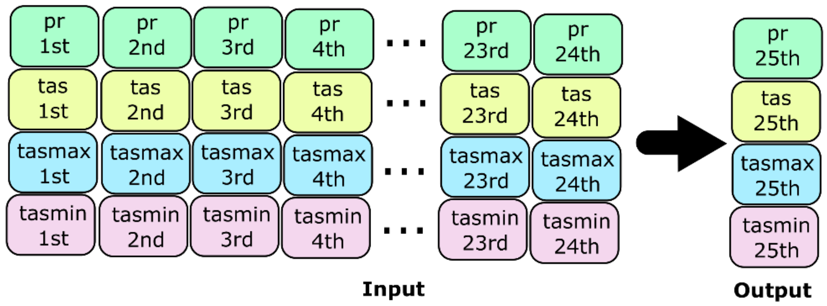

FIGURE 1. Research Predictive Scenario.

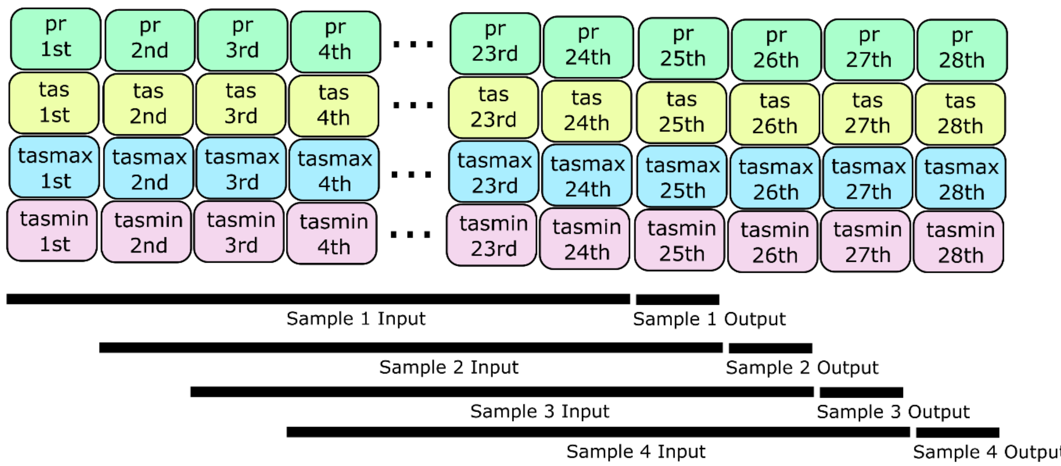

FIGURE 2. Sliding Window Implementation into Dataset.

This research used various popular ML models available on the Scikit-Learn library, such as Linear Regression as model 1, Polynomial Regression Degree 2 as model 2, Polynomial Regression Degree 3 as model 3, Decision Tree Regression as model 4, KNN as model 5, ElasticNet as model 6, Lasso as model 7, and Support Vector Regressor (SVR) as model 8. Due to the limitations of those ML models, this research implemented a flattening mechanism before inputting the data into ML models, as shown in Figure 3.

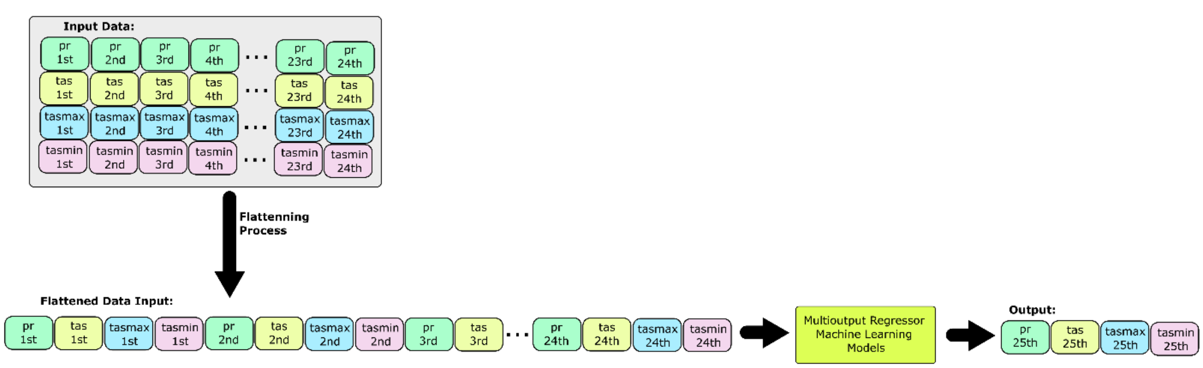

FIGURE 3. ML Process.

3.1. Dataset

The World Bank Climate Change Knowledge Portal (CCKP), which we accessed in August 2024, provided a public dataset for this research. We collected 4 data points, such as precipitation (pr), average mean surface air temperature (tas), average maximum surface air temperature (tasmax), and average minimum air temperature (tasmin), for 2 cities in Indonesia, Jakarta and Kalimantan Timur, with a detailed explanation in Table 1. The dataset is a time series dataset with recordings from 1901 January until 2022 December, or around 1464 data points, as pictured in Figure 4.

TABLE 1. Testing Result Using Jakarta Dataset.

Code	Label	Unit	Description
pr	Precipitation	mm	Precipitation accumulation.
tas	Average Mean Surface Air Temperature	°C	Average mean temperature over the aggregation period
tasmax	Average Maximum Surface Air Temperature	°C	Average daily maximum temperature over the aggregation period
tasmin	Average Minimum Surface Air Temperature	°C	Average daily minimum temperature over the aggregation period

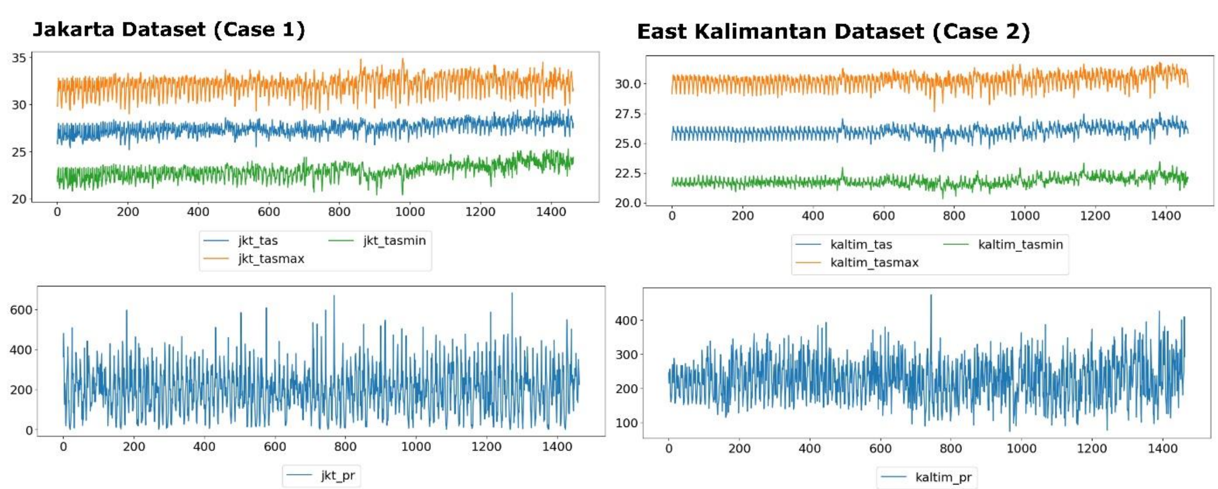

FIGURE 4. Jakarta Dataset (left) and East Kalimantan Dataset (right).

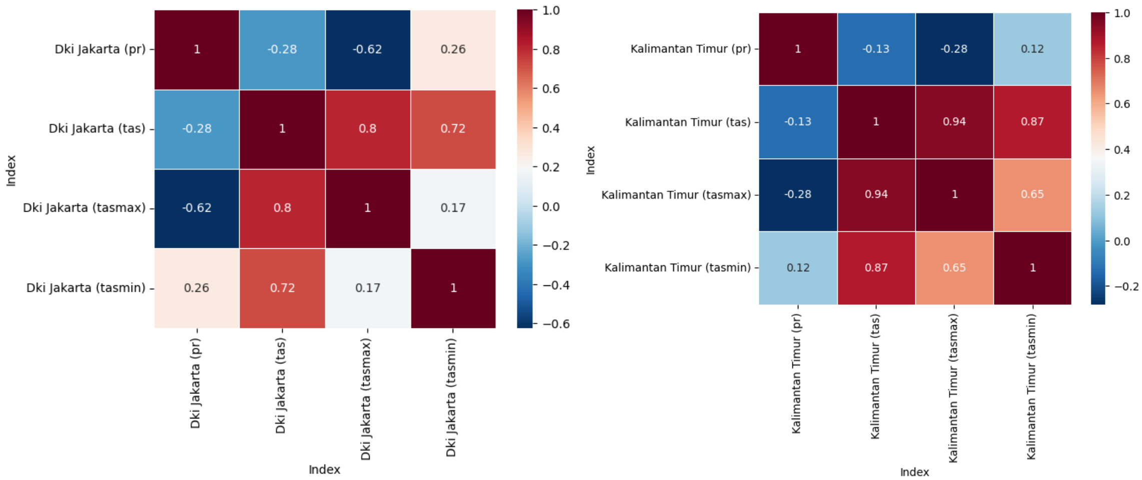

FIGURE 5. Pearson Correlation of Jakarta Dataset (left) and East Kalimantan Dataset (right).

This research uses the Pearson correlation coefficient in Figure 5 to quickly analyze the correlation between each variable. Figure 5 illustrates that the negligible correlation lies in the range of 0.00-0.10, the weak correlation in 0.10-0.39, the moderate correlation in 0.40-0.69, the strong correlation in 0.70-0.89, and the very strong correlation in 0.90-1.00, with negative values indicating the opposite direction. The Jakarta dataset has a weak, moderate, and strong correlation, while East Kalimantan has a weak, moderate, strong, and very strong correlation [14].

3.2. Linear Regression

Like most statistical methods, linear regression analysis is the simplest and most prominent approach to measuring the relationship between continuous variables, and it performs best when understanding data using practical samples [15]. Regression analysis relates a single dependent variable or predictor variable to a single independent variable or response variable can be fitted into a straight line using simple linear regression as equation 1, where y is the target/label/response/output, x is the feature/predictor/input, the C notation represents the intercept with a constant number, the multiplier (β) can be called the coefficient or weight, and ε can be mentioned as a noise.

$$y = C + \beta x + \varepsilon \tag{1}$$

Equation 2 illustrates how the ML approach can generalize linear regression to fit multiple data inputs with up to n features. We can refer to it as multiple regression, a remarkably potent technique that explores the mapping of multiple data input features into a response, and the outcome is relatively straightforward for human interpretation.

$$y = C + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \varepsilon \tag{2}$$

Nevertheless, a significant drawback of this multiple regression is its incapacity to handle a substantial number of predictors, which is referred to as multicollinearity. This instance of multicollinearity serves as an illustration of the curse of dimensionality.

3.3. Polynomial Regression

Polynomial regression is considered a special case of linear regression, as well as an enhanced form of linear regression modeled as a polynomial [16]. Equation 3 demonstrates the representation of a polynomial regression, where y represents the response variable and its coefficient [17].

$$y = C + \beta_1 x_1 + \cdots + \beta_n x_n + \cdots + \beta_{11} {x_1}^2 + \beta_{12} x_1 x_2 + \cdots + \beta_{1n\ldots n} x_1 {x_n}^{k-1} + \beta_{n\ldots n} {x_n}^k + \varepsilon \tag{3}$$

The polynomial regression model is widely favored due to its straightforward equation, well-established and comprehensible characteristics, ease of computer implementation, and reasonable adaptability to various shapes [16]. Polynomial functions are classified as a closed family. This means that when we perform linear adjustments to the data, the resulting polynomial model is still a polynomial model. This indicates that the polynomial models are not affected by the underlying metric. The polynomial model has some limitations, including poor interpolator and extrapolator properties, poor asymptotic properties, and a shape/degree tradeoff.

3.4. Decision Tree Regressor

A Decision Tree (DT) is constructed by recursively partitioning the instance space to create a classifier [18]. The DT algorithm utilizes various important terms, as depicted in Figure 6. The DT consists of several nodes, one of which is the root node. The root node acts as the initial point of the decision tree and does not receive any incoming edges. However, decision nodes have several outgoing edges and only one incoming edge. Leaf nodes in a decision tree represent the end of the tree and do not have any branches leading out, but they have exactly one branch leading in. Splitting is the process of separating a node into smaller sub-nodes, whereas pruning is the act of removing a sub-node.

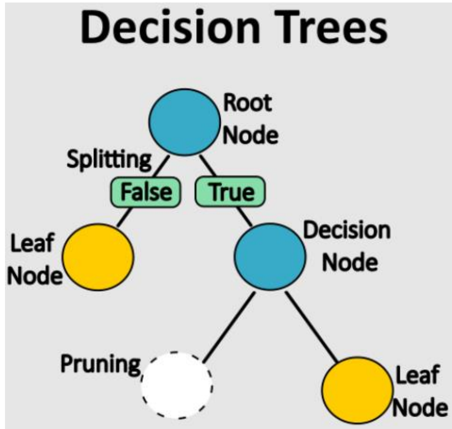

FIGURE 6. DT Structure Illustration.

In the case of the decision tree regressor, or decision tree in regression, the structure of the tree is essentially the same. However, the key distinction lies in how it predicts the value at each leaf node. Instead of using a specific value, it calculates the mean/average of the target variable values at that node. Regression evaluation metrics like mean squared error (MSE) and mean absolute error (MAE) can evaluate this prediction result.

3.5. K-Nearest Neighbor Regressor

Even though the K-Nearest Neighbor (KNN) is a popular ML model in classification tasks, the KNN-Regressor Method is a supervised ML method that can perform well on small amounts of data [19]. The technique calculates the distances between the query and all instances in the data by selecting a specific number of samples (K) that are closest to the query. It then computes the average of the labels in the case of a regression problem. Usually, we use the Euclidean distance to gauge how close a query is to other data [20].

3.6. Lasso

Often used for parameter reduction, the lasso regression model is a linear model that estimates sparse parameters [21]. For this reason, compressed sensing commonly employs the Lasso regression model. Lasso can be succinctly described as follows: Initially, we employ a general linear model represented by equation 2. Subsequently, we determine the optimal fit and strive to minimize the residual sum of squares (RSS) [21]. RSS, or Residual Sum of Squares, is the quantitative measure of the discrepancy between the observed value and the projected value. Regularization can be incorporated into the RSS minimization process by introducing a contraction penalty term. This penalty term includes the normalized values of the α and β coefficients and weight.

3.7. ElasticNet Regression

The ElasticNet Regression was developed in response to criticism of the Lasso regression model, which indicated that the Lasso model could become overly reliant on the data, resulting in instability [22]. ElasticNet proposed combining the penalties of Ridge and Lasso regression to achieve optimal results. In brief explanation, ElasticNet regression is a regularization technique that combines the penalties of Lasso and Ridge regression methods. It is especially beneficial when dealing with many correlated variables.

3.8. Support Vector Regressor (SVR)

Support Vector Regression (SVR) is a variant of the Support Vector Machine (SVM) classification technique. It solves an optimization problem to develop a regression function that maps predictor variables to observed target values [23]. Support Vector Regression (SVR) is highly effective for high-dimensional data. In comparison to other regression models, SVR offers distinct advantages. It efficiently handles nonlinear regression problems by projecting the original feature into a kernel space, enabling linear discrimination of the data. Additionally, SVR can learn a model that describes the importance of a variable and characterizes the relationship between features and the target variable.

4. ANALYSIS AND RESULTS OF EXPERIMENTAL RESULTS

For standard practice in ML research, this research split the dataset into two parts: training with 80% blue-colored data and testing with 20% orange-colored data, as shown in Figure 7. Figure 7 shows the time series data in Jakarta region on the right, whereas East Kalimantan/Kalimantan Timur is on the left. The training data began in January 1901 until July 1996, and the testing data began in August 1996 until December 2022 for both scenarios in the two regions of Jakarta and East Kalimantan.

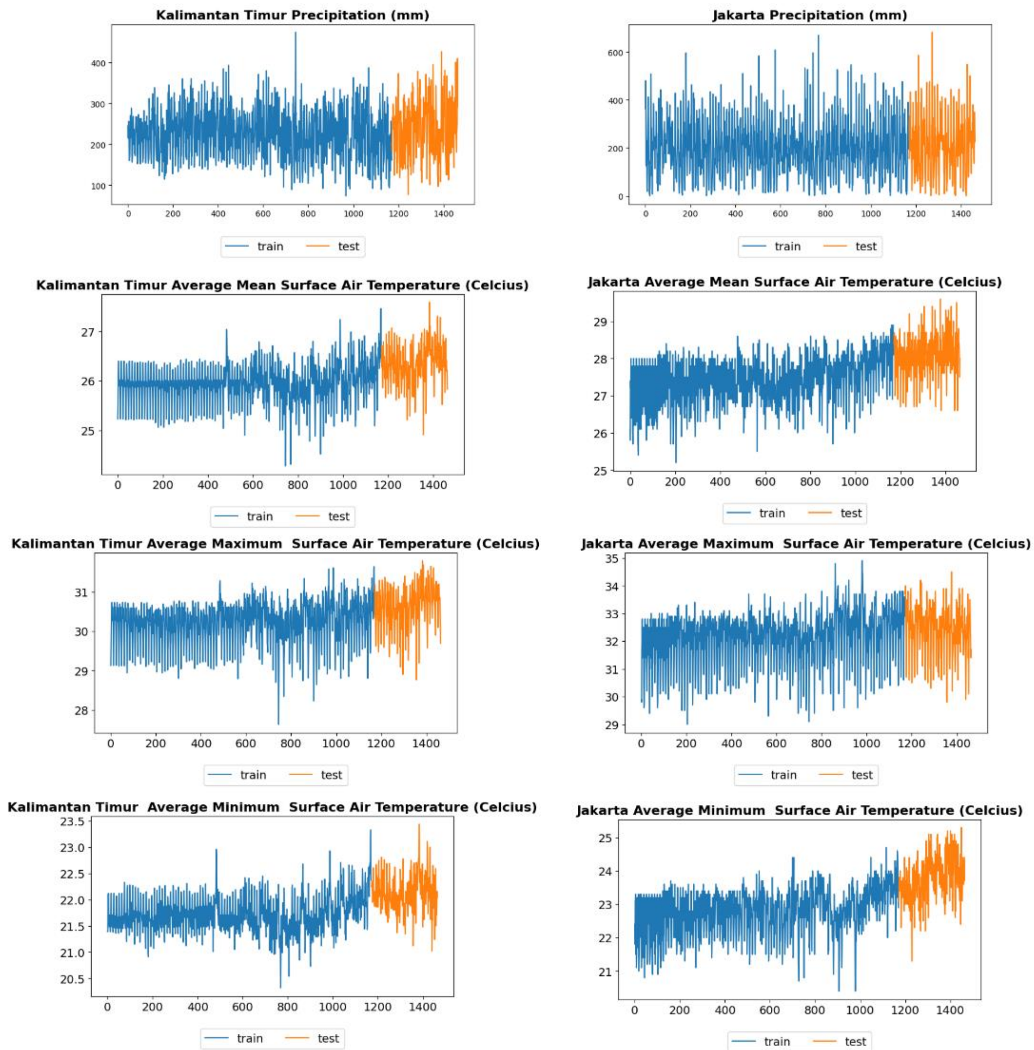

FIGURE 7. Dataset Split of Jakarta Dataset (right) and East Kalimantan Dataset (left).

TABLE 2. ML Testing Result Using Jakarta Dataset.

TABLE 2. ML Testing Result Using Jakarta Dataset.

ML Model	pr (mm)		tas (°C)		tasmax (°C)		tasmin (°C)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Model 1: Linear Regression	75.4812	94.7021	**0.3821**	**0.4737**	**0.4564**	**0.5799**	**0.4650**	**0.6065**
Model 2: Polynomial Regression Degree 2	211.8050	284.9102	1.2375	1.6195	1.4477	1.8300	1.2983	1.6729
Model 3: Polynomial Regression Degree 3	164.4217	213.8839	0.8724	1.1691	1.0096	1.2916	1.0583	1.4358
Model 4: Decision Tree Regression	95.3164	124.6392	0.5901	0.7601	0.6881	0.8941	0.6642	0.8466
Model 5: KNN Regression	**66.6880**	**84.5756**	0.6885	0.8483	0.5349	0.6776	1.0973	1.2946
Model 6: ElasticNet	69.1537	86.6382	0.7317	0.8711	0.5866	0.7417	1.0976	1.2858
Model 7: Lasso	70.1314	87.8159	0.7308	0.8675	0.5841	0.7378	1.0954	1.2811
Model 8: SVR	79.4403	101.5636	0.7051	0.8594	0.5629	0.7117	1.0824	1.2757

TABLE 3. ML Testing Result Using East Kalimantan Dataset.

ML Model	pr (mm)		tas (°C)		tasmax (°C)		tasmin (°C)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Model 1: Linear Regression	40.2540	50.6576	**0.2080**	**0.2657**	**0.2533**	**0.3230**	**0.2040**	**0.2674**
Model 2: Polynomial Regression Degree 2	123.1719	153.6251	0.4279	0.5533	0.5710	0.7372	0.5270	0.6714
Model 3: Polynomial Regression Degree 3	96.1453	123.4166	0.4198	0.5301	0.5802	0.7705	0.4857	0.6405
Model 4: Decision Tree Regression	59.1040	75.7750	0.3294	0.4230	0.4200	0.5416	0.3105	0.3967
Model 5: KNN Regression	40.9381	52.2374	0.4977	0.5656	0.5359	0.6179	0.4832	0.5503
Model 6: ElasticNet	40.0071	51.0643	0.4637	0.5349	0.5049	0.5937	0.4473	0.5146
Model 7: Lasso	**39.6637**	**50.5797**	0.4605	0.5323	0.5012	0.5882	0.4457	0.5139
Model 8: SVR	50.2935	62.3915	0.4566	0.5274	0.5178	0.6065	0.4156	0.4820

Table 2 displays the ML testing comparison results for Jakarta data. According to Table 2, we can conclude that the data was quite simple because the implementation of a linear regression model outperformed all ML models in comparison. When this study tried to increase the complexity of the linear regression model by implementing a polynomial approach to increase the dimensionality of the input features, the results showed that the overall error in MAE and RMSE increased significantly. Overall, the best ML model in this research was linear regression because it can cause the smallest error in predicting tas, $tasmax$, and $tasmin$ variables, but it is still good enough in predicting pr variables to be very competitive with the second-best model, Elastic-Net. The worst model in this Jakarta scenario was polynomial regression with a second degree.

Table 3 shows the ML testing comparison result for East Kalimantan data. A quick glance at Table 3 reveals that the error trend of each ML model is not significantly different from Table 2, where linear regression emerged as the best model for the tas, tasmax, and tasmin variables. However, when it came to predicting PR variables, it was still highly competitive with the Lasso model. It seems that the East Kalimantan climate variables were not very complex due to the improvement of linear regression. Implementing a polynomial approach to increase the dimensionality of the dataset can cause the model to become worse or increase the error in MAE and RMSE.

To enhance comprehension of the research results, Figures 8, 9, 10, 11, 12, 13, 14 and 15 illustrate the testing results of each variable in each ML model for both Jakarta and East Kalimantan data. We achieve this by comparing the ground truth data, represented by blue lines, with the predicted results, represented by orange lines. It can be seen from Figure 8 that linear regression and elastic net follow the pr testing variable pattern better than other models, which closely match the results of Table 2. Meanwhile, the polynomial approach seems like too many errors.

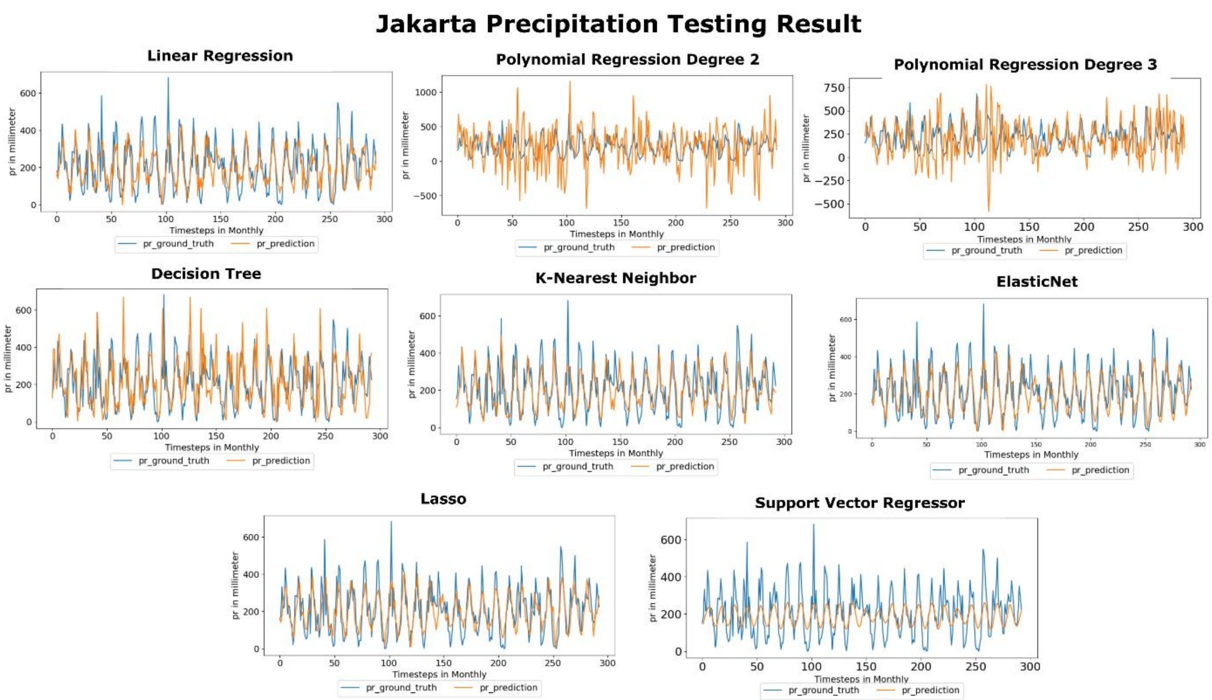

FIGURE 8. Comparison of ML Models in Predicting Jakarta Precipitation (pr).

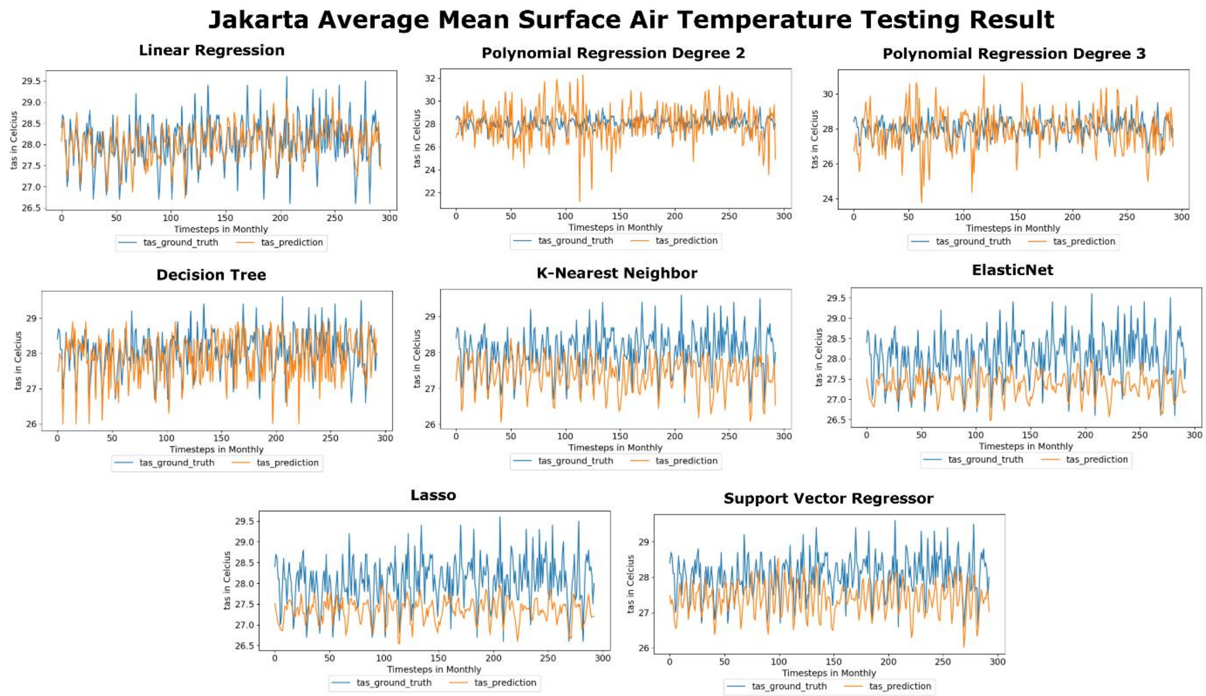

FIGURE 9. Comparison of ML Models in Predicting Jakarta Average Mean Surface (tas).

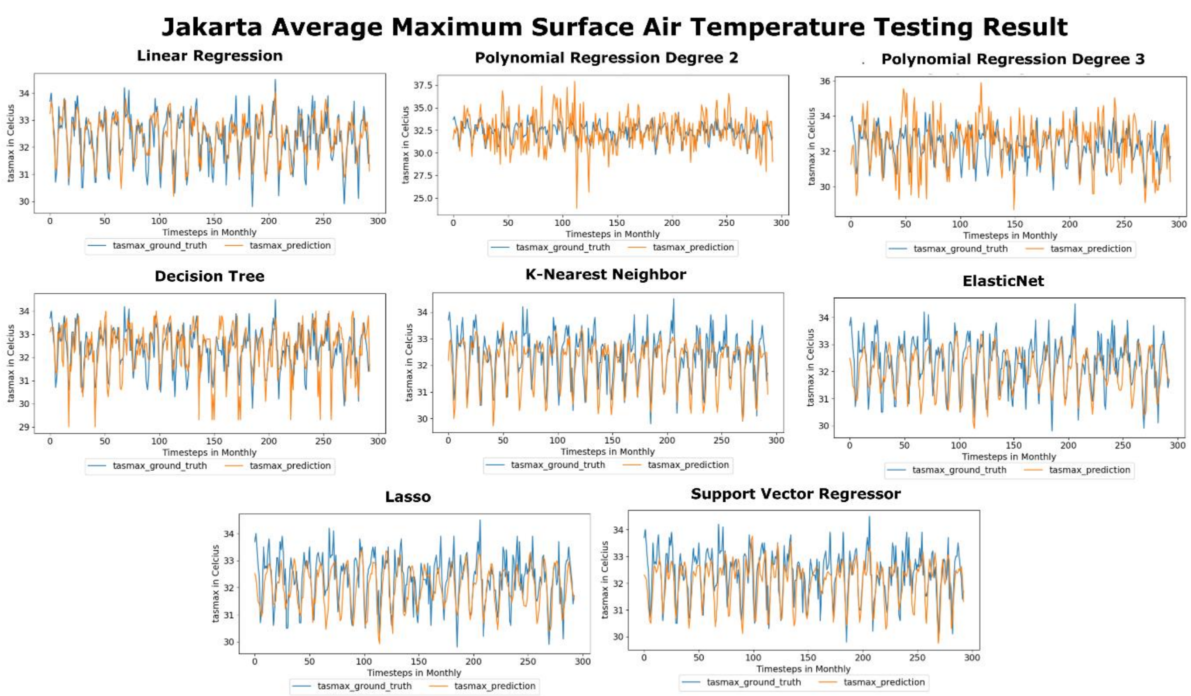

FIGURE 10. Comparison of ML Models in Predicting Jakarta Average Maximum Surface (tasmax).

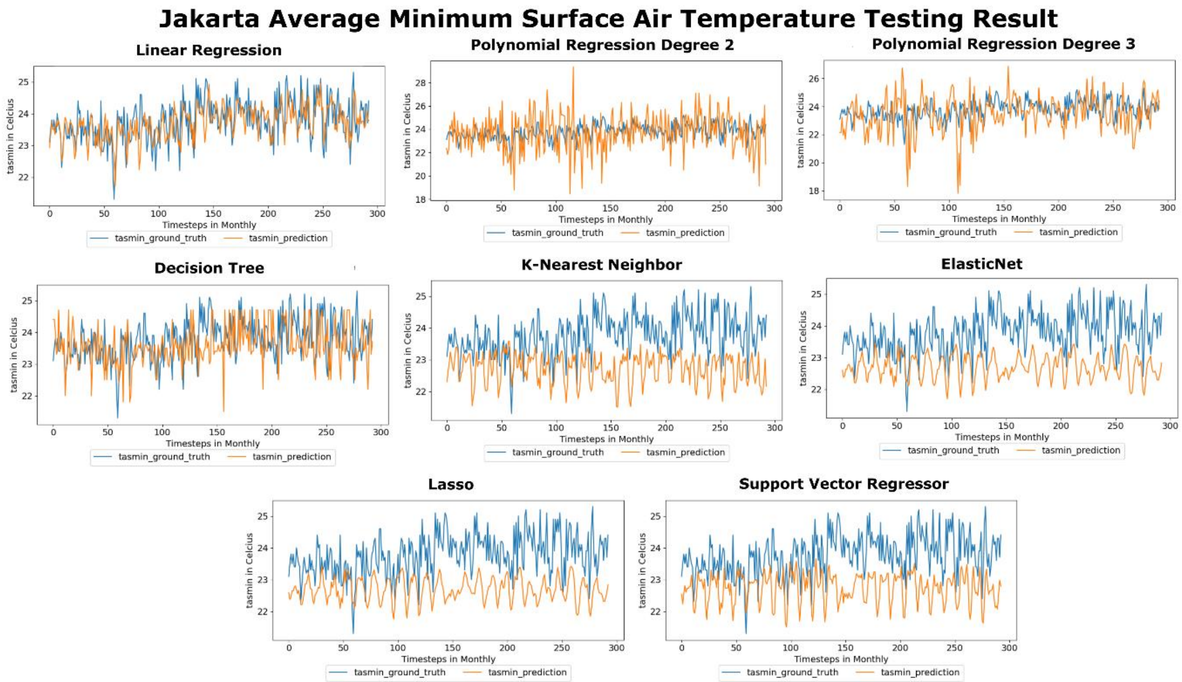

FIGURE 11. Comparison of ML Models in Predicting Jakarta Average Minimum Surface (tasmin).

Overall, according to Figures 8, 9, 10 and 11, the visualization of the linear regression testing result appears to follow the pattern of testing data in all four climate condition variables, as shown in Table 2. The polynomial approach yielded excessive error due to its sharp fluctuations, making it the least effective model. We can conclude that both polynomials with 2- and 3-degrees, as well as the decision tree model, tend to overfit and underfit in certain details. All models seem to struggle to predict extreme values at peak points and steep valleys.

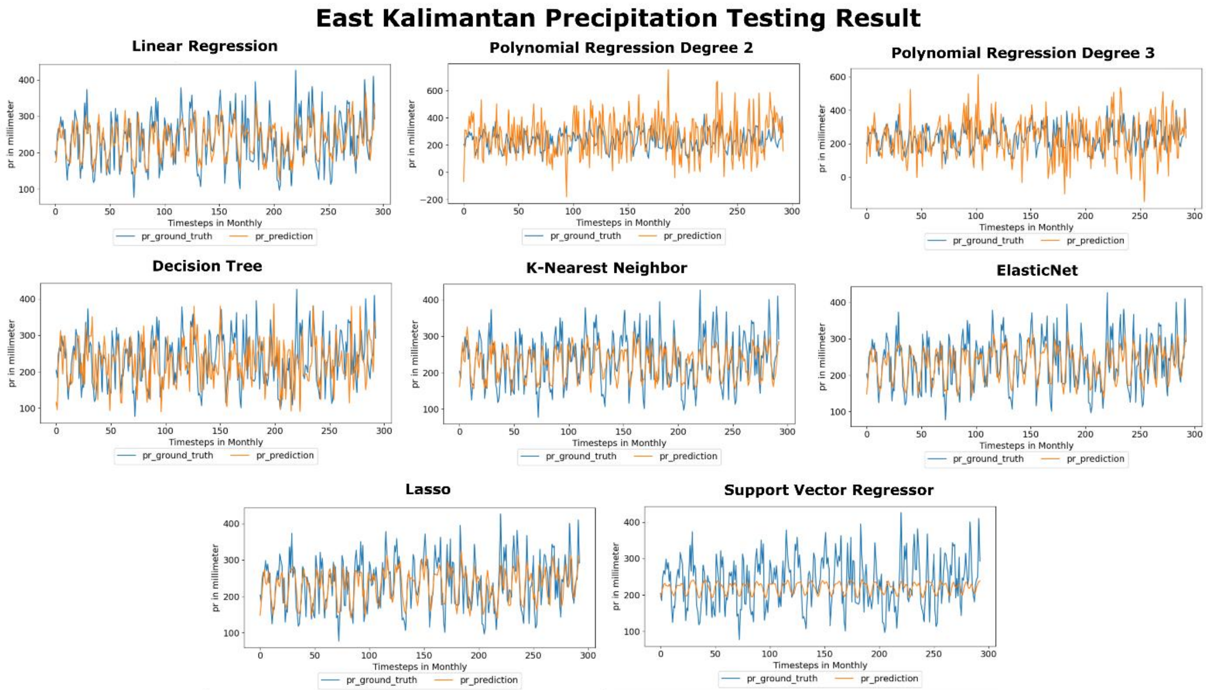

FIGURE 12. Comparison of ML Models in Predicting East Kalimantan Precipitation (pr).

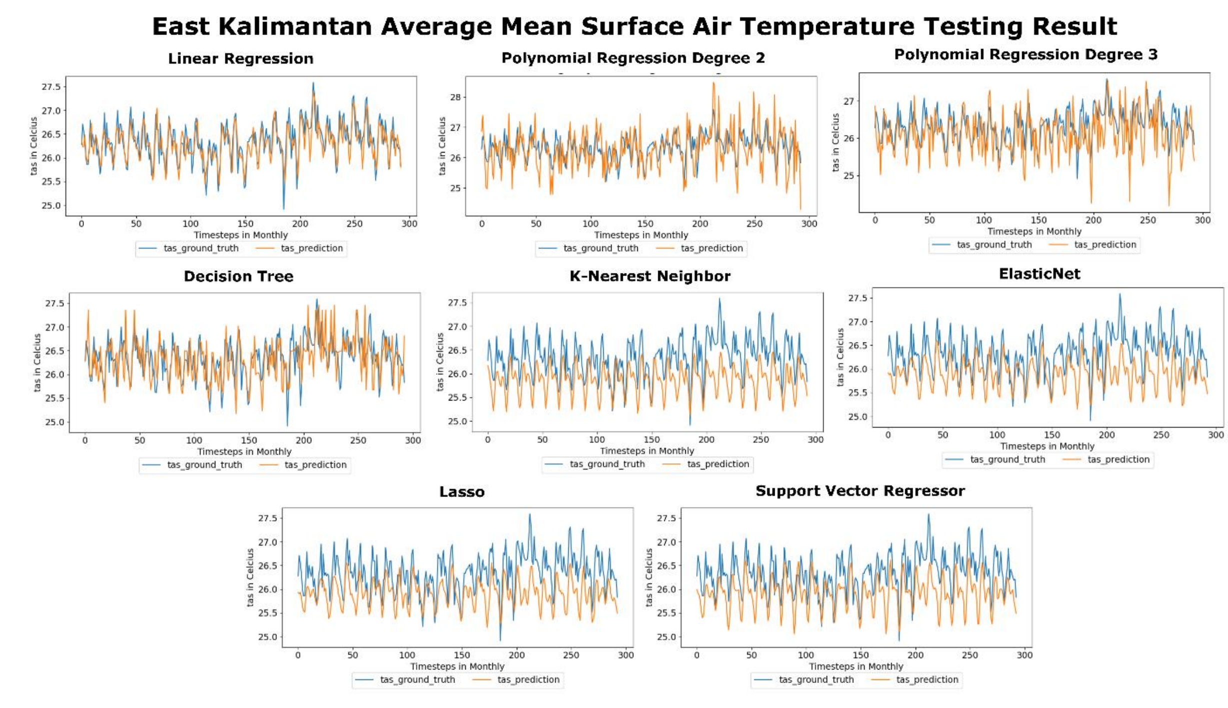

FIGURE 13. Comparison of ML Models in Predicting East Kalimantan Average Mean Surface (tas).

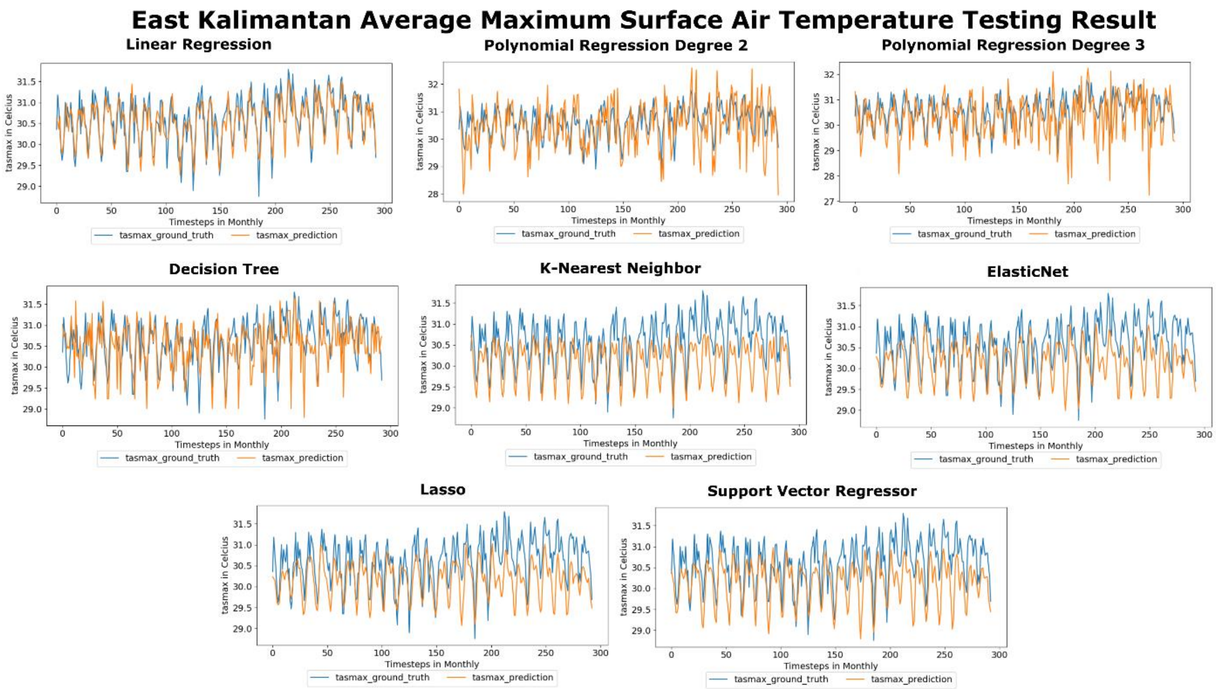

FIGURE 14. Comparison of ML Models in Predicting East Kalimantan Average Maximum Surface (tasmax).

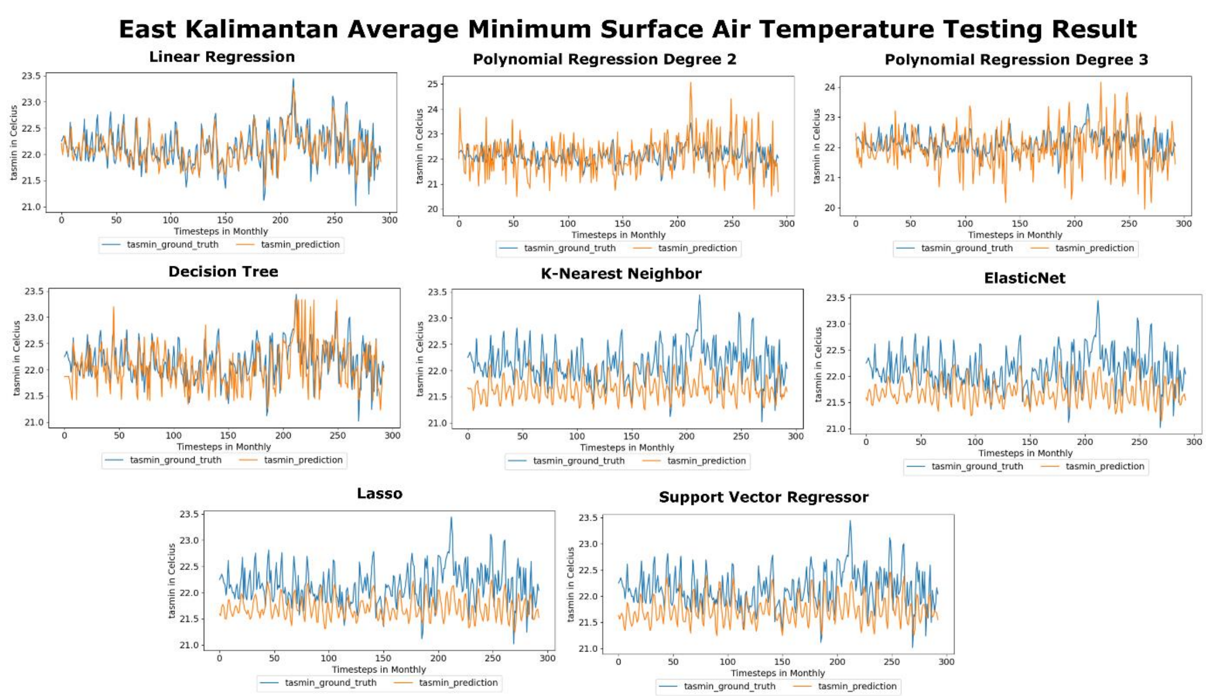

FIGURE 15. Comparison of ML Models in Predicting East Kalimantan Average Minimum Surface (tasmin).

Figures 12, 13, 14 and 15 depict the overall results of comparison testing on East Kalimantan data using all ML models. The trend is not much different compared to Jakarta data testing. Certain models, including polynomial regression with 2- and 3-degrees and decision trees, also exhibit overfitting and underfitting in specific areas. Some models, such as linear regression, KNN, elastic-net, Lasso, and SVR, seem to follow the pattern of East Kalimantan data testing.

Other ML architectures, such as ensemble learning, which combine various base ML models, may increase performance and reduce error for future work. Predicting future climate conditions may be a complex task; future work may combine other environmental variables, which may provide valuable insight [24]. This future work should further explore promising deep learning approaches such as recurrent neural networks, gated recurrent units, long short-term memory, seq2seq, and others [25].

5. CONCLUSION

This study developed eight different ML models, such as linear regression, polynomial regression with 2- and 3-degrees, decision tree regressor, KNN regressor, elastic-net, Lasso, and SVR, to forecast the climate condition variables in two different regions in Indonesia, such as Jakarta, the previous capital city, and East Kalimantan, which represents the location of the newest Indonesian capital city. The results show that the datasets in both Jakarta and East Kalimantan are quite simple to understand with a simple ML model such as linear regression. The best ML model for both regions was linear regression. The other models, such as KNN, elastic-net, lasso, and SVR, were not bad because they could follow the pattern of data testing ground truth; meanwhile, models like polynomial regression with 2- and 3-degrees and decision tree regressors appear to be overfit and underfit for certain spots.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] R. Kalalinggi, M. Hisdar, M. Sarmiasih, et al. Forecasting the development of IKN (New National Capital) in sustainable development, Indonesia, J. Govern. Public Policy. 10 (2023), 48-59. https://doi.org/10.18196/jgpp.v10i1.16786.

[2] P. Van de Vuurst, L.E. Escobar, Perspective: Climate change and the relocation of Indonesia's capital to Borneo, Front. Earth Sci. 8 (2020), 5. https://doi.org/10.3389/feart.2020.00005.

[3] D.L. Corwin, Climate change impacts on soil salinity in agricultural areas, Eur. J. Soil Sci. 72 (2020), 842–862. https://doi.org/10.1111/ejss.13010.

[4] M. Fathi, M.H. Kashani, S.M. Jameii, et al. Big data analytics in weather forecasting: A systematic review, Arch. Comput. Methods Eng. 29 (2021), 1247–1275. https://doi.org/10.1007/s11831-021-09616-4.

[5] D. Scott, Sustainable tourism and the grand challenge of climate change, Sustainability 13 (2021), 1966. https://doi.org/10.3390/su13041966.

[6] S. Fathi, R. Srinivasan, A. Fenner, et al. Machine learning applications in urban building energy performance forecasting: A systematic review, Renew. Sustain. Energy Rev. 133 (2020), 110287. https://doi.org/10.1016/j.rser.2020.110287.

[7] B. Bochenek, Z. Ustrnul, Machine learning in weather prediction and climate analyses—Applications and perspectives, Atmosphere 13 (2022), 180. https://doi.org/10.3390/atmos13020180.

[8] C.O. de Burgh-Day, T. Leeuwenburg, Machine learning for numerical weather and climate modelling: a review, Geosci. Model Dev. 16 (2023), 6433–6477. https://doi.org/10.5194/gmd-16-6433-2023.

[9] K. Ahmed, D.A. Sachindra, S. Shahid, et al. Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms, Atmos. Res. 236 (2020), 104806. https://doi.org/10.1016/j.atmosres.2019.104806.

[10] Y. Zhang, A. Ye, Machine learning for precipitation forecasts post-processing — Multi-model comparison and experimental investigation, J. Hydrometeorol. 22 (2021), 3065–3085. https://doi.org/10.1175/jhm-d-21-0096.1.

[11] J. Cifuentes, G. Marulanda, A. Bello, et al. Air temperature forecasting using machine learning techniques: A review, Energies. 13 (2020), 4215. https://doi.org/10.3390/en13164215.

[12] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Comparison of deep learning sequence-to-sequence models in predicting indoor temperature and humidity in solar dryer dome, Commun. Math. Biol. Neurosci. 2022 (2022), 98. https://doi.org/10.28919/cmbn/7655.

[13] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Indoor climate prediction using attention-based sequence-to-sequence neural network, Civ. Eng. J. 9 (2023), 1105–1120. https://doi.org/10.28991/cej-2023-09-05-06.

[14] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: Appropriate use and interpretation, Anesth. Analg. 126 (2018), 1763–1768. https://doi.org/10.1213/ane.0000000000002864.

[15] T.M.H. Hope, Linear regression, in: Machine Learning, Elsevier, 2020: pp. 67–81. https://doi.org/10.1016/B978-0-12-815739-8.00004-3.

[16] A. Peckov, A machine learning approach to polynomial regression, PhD thesis, Jozef Stefan International Postgraduate School, Ljubljana, 2012.

[17] P. Morala, J.A. Cifuentes, R.E. Lillo, et al. Towards a mathematical framework to inform neural network

modelling via polynomial regression, Neural Networks. 142 (2021), 57–72. https://doi.org/10.1016/j.neunet.2021.04.036.

[18] K.E. Setiawan, Predicting recurrence in differentiated thyroid cancer: a comparative analysis of various machine learning models including ensemble methods with chi-squared feature selection, Commun. Math. Biol. Neurosci. 2024 (2024), 55. https://doi.org/10.28919/cmbn/8506.

[19] A.N. Beskopylny, S.A. Stel'makh, E.M. Shcherban', et al. Concrete strength prediction using machine learning methods CatBoost, k-nearest neighbors, support vector regression, Appl. Sci. 12 (2022), 10864. https://doi.org/10.3390/app122110864.

[20] S. Shabani, S. Samadianfard, M.T. Sattari, et al. Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis, Atmosphere 11 (2020), 66. https://doi.org/10.3390/atmos11010066.

[21] S. Wang, Y. Chen, Z. Cui, et al. Diabetes risk analysis based on machine learning LASSO regression model, J. Theory Pract. Eng. Sci. 4 (2024), 58-64.

[22] P.J. García-Nieto, E. García-Gonzalo, J.P. Paredes-Sánchez, Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques, Neural Comput. Appl. 33 (2021), 17131–17145. https://doi.org/10.1007/s00521-021-06304-z.

[23] F. Zhang, L.J. O'Donnell, Support vector regression, in: Machine Learning, Elsevier, 2020: pp. 123–140. https://doi.org/10.1016/B978-0-12-815739-8.00007-9.

[24] R. Fredyan, K.E. Setiawan, An investigation into improving El Niño-Southern Oscillation prediction based on temporal transformer architecture, Commun. Math. Biol. Neurosci. 2024 (2024), 42. https://doi.org/10.28919/cmbn/8371.

[25] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Sequence to sequence deep learning architecture for forecasting temperature and humidity inside closed space, in: 2022 10th International Conference on Cyber and IT Service Management (CITSM), IEEE, Yogyakarta, Indonesia, 2022: pp. 1–7. https://doi.org/10.1109/CITSM56380.2022.9936008.