Contents lists available at JESA

Journal of Engineering and Science Application

journal homepage: https://jesa.aks.or.id/index.php/jesa/index

Evaluating Imputation Approaches and Support Vector Regression Parameters in Weather Forecasting

Arif Mudi Priyatno[a], Yunia Ningsih[b]
[a]*Universitas Pahlawan Tuanku Tambusai, Riau, Indonesia*
[b]*Universitas Trisakti, Jakarta, Indonesia*

Article Info

Keywords:

Rainfall;
Data Imputation;
Support Vector Regression;
K-Nearest Neighbor;
Radial Basis Function Kernel;
Artificial Neural Network.

ABSTRACT

Rainfall plays a vital role in various sectors such as transportation, agriculture, and industry. Having accurate rainfall information enables stakeholders in these fields to take proper measures and minimize potential losses caused by inaccurate data. This study focuses on identifying an effective method for rainfall forecasting by examining imputation techniques in data preprocessing and parameter settings within Support Vector Regression (SVR). The experimental findings indicate that the most effective imputation method for SVR is determined using the Mean Squared Error (MSE) and Mean Absolute Error (MAE) evaluation metrics. Based on MSE, the k-nearest neighbor method proves to be the most reliable approach for data imputation preprocessing. The preprocessing results were then applied to Polynomial SVR with parameters C = 1000, tolerance = 0.001, epsilon = 0.01, and unlimited iterations. Conversely, MAE results highlight Artificial Neural Network (ANN) as the optimal imputation method. ANN, when combined with a radial basis function kernel, gamma = 0.001, C = 1000, tolerance = 0.001, and unlimited iterations, was further tested using RBF SVR under the same parameter settings.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Arif Mudi Priyatno
Universitas Pahlawan Tuanku Tambsuai
Email: arifmudi@universitaspahlawan.ac.id

1. INTRODUCTION

Uncertainty is a common problem in many natural sciences, such as meteorology and hydrology. This makes predicting natural conditions, such as rainfall, a worthy subject for study by each of these sciences [1]. Rainfall is the most important factor related to the water cycle. Rainfall characteristics are crucial for modeling the accuracy and predictability of other hydrological values, such as runoff and evapotranspiration. This is because rainfall characteristics control modeling behavior and evaluation [2] [3] [4].

Weather is the condition of the atmosphere at a location over a relatively short period of time. Parameters influencing weather conditions in an area include temperature, air pressure, wind speed, humidity, and various other atmospheric phenomena [5] [6] [7]. Selecting the right method for determining weather conditions is a frequently undertaken activity by several weather or atmospheric researchers in recent

years. This is due to the growing demand from various parties for faster, more accurate, and more detailed information on atmospheric conditions [8] [9].

For rainfall forecasts, data is obtained from weather agencies. Meteorological and geophysical agencies generally provide the required data, accessible through their portals. The data presents the values of parameters that influence rainfall. Data taken from a live event process is referred to as real data. Rainfall is included in real data that has a large size. However, the use of real data will result in data anomalies. Anomalous data is usually caused by various factors such as sensor detection factors and human error itself. Anomalous data will certainly affect the results of rainfall forecasting. Therefore, improvements to the data before entering the forecasting process are necessary. This paper aims to analyze the comparison of imputation models and the effect of changes in parameter values in the Support Vector Regression (SRV) method for rainfall forecasting. Imputation data is obtained from the data preprocessing process using three tested algorithms. The algorithms used to generate imputed data are the K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Binning. Meanwhile, the SVR parameters analyzed are gamma, C parameter, tolerance, epsilon, and the number of iterations. The imputation data model with changes in the SVR method parameter values is combined to obtain the best preprocessing algorithm and parameter values for rainfall forecasting. The best algorithm and parameter values are determined based on the MSE and MAE values of each experimental technique. This research gained knowledge about preprocessing algorithms for addressing anomalous data. Furthermore, it also provided insight into the impact of changing parameter values on the SVR method in the classification process.

2. RESEARCH METHOD

This study used primary data from the Sultan Syarif Kasim II Meteorological Station in Pekanbaru, Riau, from January 1, 2000, to October 13, 2018 [10]. The data contained 10 features: minimum temperature (°C), maximum temperature (°C), average temperature (°C), average humidity (%), rainfall (mm), irradiation time (hours), average wind speed (knots), highest wind direction (deg), highest wind speed (knots), and wind direction at maximum speed (deg). The data contained anomalous values, as there were data values of "8888" and "9999." A value of "8888" indicates that the weather parameter value was not measured on a given day, while a value of "9999" indicates that the data value was not recorded for one day. The total number of data obtained was 6,851.

This study implemented several methods to improve anomalous data. This process is a preprocessing of the used data. The preprocessing step produced three imputed data based on the used algorithm. The algorithms used were the k-nearest neighbor algorithm, Artificial Neural Network, and binning. After preprocessing was completed, each imputed data was generated for the outlier process. In this process, the data was corrected using the standard deviation method to remove outliers. Three pieces of corrected imputed data were then fed into the forecasting process using the Support Vector Regression method.

2.1 Outlier Process

Outliers are observations that are far (extreme) from other observations. Generally, outliers are divided into two categories: outliers in observational models and outliers in linear models. Based on the number of variables considered, outliers are divided into univariate and multivariate observational outliers. Outliers in multivariate linear models can be divided into three categories: outliers for leverage and residuals, or both [11][12]. Outlier detection is a common process in data mining applications. Detection is performed to avoid distributions between data with significantly different values. Outlier detection in data mining is often based on distance measures, clustering, and spatial methods [13]. In outlier detection, data in data mining processing can be categorized into univariate and multivariate data [14][15]. Univariate data is data that contains a single type of variable. Multivariate data, on the other hand, contains more than one type of data. In this study, the data used was data with a multivariate model.

2.1.1 K-Nearest Neighbors Algorithm (KNN)

KNN is included in the Lazy Learning category. The KNN method approach is carried out by storing all training data in an n-dimensional space. When testing data, it will determine the K nearest neighbors which will be the values of the testing data. The KNN method is used to correct anomalous data. The algorithm produces imputed data ready to enter the forecasting process. The rainfall data used in this study is characterized into two based on the completeness of the values of each parameter. First, the data for each row has complete parameter values. There are no missing values for any parameter. Second, the data for each row has missing values for a parameter. In the data used, missing values are marked with the values "8888" and "9999". Data with the first characteristic becomes clean data or training data. Meanwhile, data

with the second characteristic becomes unclean data or test data that will be corrected. Algorithm 1 is an algorithm for dealing with unclean data.

Algorithm 1: Replacing data containing missing values using the KNN method

Input: *Clean, unclean data*
Output: *Matrix data resulting from missing value correction using the KNN method*
Steps:

a. Check each row for "unclean" data.
b. Check the number of missing values per row.
If the number of missing values per row is more than five, the row is deleted and step 1 is repeated. Otherwise, proceed to the next step.
c. Check the data in features containing missing values. Data in features that do not contain missing values will become the dataset, and data in features that contain missing values will be used as the target data.
d. KNN training uses clean data, provided that the dataset and target data match those obtained in step 3.
e. Prediction is performed on the tested data rows.
f. The results of predictions that do not contain missing values are added to the clean data and become new data.
g. The new data is stored in a matrix, and step 1 is repeated until all "unclean" data has been processed into clean data.

2.1.2 Artificial Neural Network Algorithm (ANN)

In general, the data repair process using the ANN method is similar to the KNN method. However, in repairing each row of data, this method creates network parameters that are used to determine missing values in the corrected data. Steps 4 and 5 explain how the data repair process using the ANN method produces imputed data.

Algorithm 2: Replacing data containing missing values using the Artificial neural network method

Input: Clean data, unclean data
Output: Matrix data resulting from missing value correction using the Artificial Neural Network method
Steps:

a. Check the number of missing values per row.
b. If the number of missing values per row is more than five, the row is deleted and step 1 is repeated. Otherwise, proceed to the next step.
c. Check the data in features containing missing values. Data in features that do not contain missing values will become the dataset, and data in features containing missing values will be used as the target data.
d. ANN training uses clean data, provided that the dataset and target data match those obtained in step 3.
e. Prediction is performed on the tested data rows.
f. The results of predictions that do not contain missing values are added to the clean data and become new data.
g. The new data is stored in the matrix, and step 1 is repeated until all unclean data has been processed into clean data.

2.1.3 Binning Algorithm

The binning method is proposed to improve data by sorting neighboring values of that value. The sorting data is distributed into several 'bins'. Because the data distribution in the binning method considers neighboring values, the resulting value in this binning method is the average value of the bin. Calculations in the binning method commonly used are to calculate the average in each "bin" or calculate the median in each "bin". The binning method in this study is used as a third method to improve the missing value parameters. Calculations in the binning method are used by calculating the middle value. Step 2 is the algorithm for improving data with this method.

Algorithm 3: Replacing data containing missing values using the binning method

Input: All Data
Output: Data corrected using the binning method
Steps:

a. Check the number of missing values per row. If the number of missing values per row is more than five, the row is deleted and step 1 is repeated. Otherwise, proceed to the next step.
b. Check the data in the features containing missing values. Each row will calculate the median value in that feature to replace the missing value.
c. Repeat the second step for as many rows as the data passed in step 1, and each row is repeated for as many features as possible.
d. The corrected data for missing values is stored in the clean data.

2.2 Support Vector Regression

Support Vector Regression (SVR) is one method for solving problems in regression. SVR is used to find a function g(x) that fits all data where the value of e approaches the agreed value or reaches the specified iteration. The value of e indicates the accuracy of the SVR calculation. SVR consists of several parameters, namely kernel, degree, gamma, coef0, tolerance, parameter, epsilon, shrink, cache_size, verbose, and maximum iteration. However, the testing of the imputation data model in this study did not use all SVR parameters. The parameters used are kernel, gamma, tolerance, epsilon, variable C, and maximum iteration. The SVR kernels used for testing are RBF, Linear, and Polynomial kernels. While for the other parameter values, they are with the specified values. Table 1 is a description of the parameters and their values for testing the imputation data.

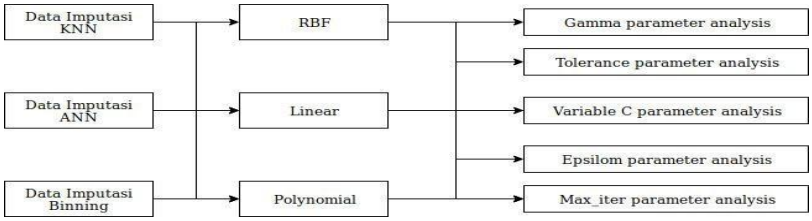


Figure 1. Imputation Data Testing Model Using the SVR Method

Table 1. SVR test parameters and their values

Kernel	***RBF, Linear, Polynomial***				
Gamma	1	10	100	1000	10000
tolerance	0.1	0.01	0.001	0.0001	0.00001
Variable C	1	1.00E+02	1.00E+03	1.00E+04	1.00E+05
Epsilon	0.1	0.01	0.001	0.0001	0.00001
Maximum Iteration	50	100	500	1000	-1

3. RESULTS AND DISCUSSION

This section describes the results of comparative research on imputation data models and parameter value changes for rainfall forecasting. The results of the imputation data testing classified by the SVR method are presented in a test results table. The optimality level of the testing technique is displayed in the form of two parameter values: MSE and MAE. Figure 1 illustrates the imputation data testing process obtained using three preprocessing methods. The test results will be explained in the parameter analysis subsection. There are five parameter analysis subsections, as shown in Figure 1. This analysis concluded that a good data enhancement model was used, along with optimal kernel values and parameters.

3.1 Gamma Parameter Analysis

The gamma parameter is the inverse of the standard deviation of the RBF kernel (Gaussian function), which is used as a measure of similarity between two points. In SVM with RBF kernel, gamma is used to accelerate the functions in the RBF kernel to obtain optimal classification model accuracy. Figure 2 is a graph of the MSE and MAE results from the imputation data testing. The results are illustrated in graphical form because there are 3 imputation data models tested. The gamma parameter is only owned by the RBF kernel. Therefore, each imputation data model has a single graph value line. The smallest MSE value is 0.012

in the KNN imputation data located at the gamma parameter value of 0.3 kernel. While the MAE value is 0.064 in the ANN imputation data that is split at the gamma parameter value of 0.001.

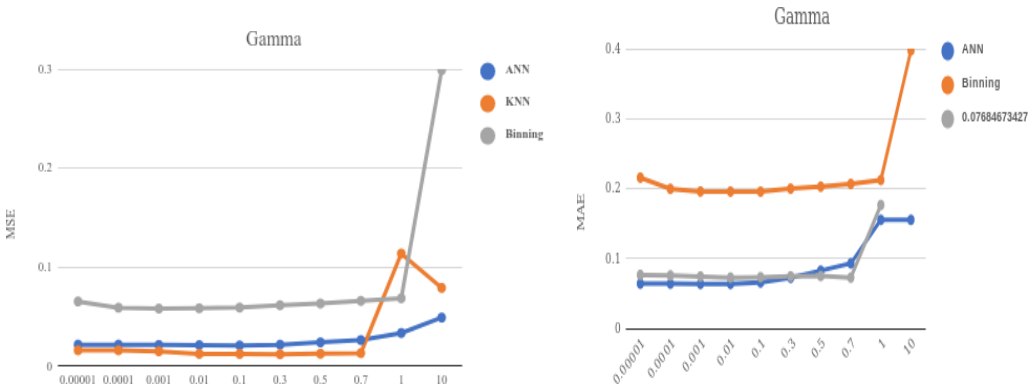


Figure 2. MSE and MAE Graphs of Changes in Gamma Parameter Values

3.2 Tolerance Parameter Analysis

Tolerance is used to determine the lowest tolerance value in the specified SVR process. In general, the smaller the tolerance value, the better the value obtained. Table 2 shows that if the smallest tolerance value has been found, the next one will simply follow the previous tolerance and the changes will not change significantly. The lowest MSE value is generated from the polynomial kernel with the KNN preprocessing method. The MSE value is 0.013, while the minimum MAE produced is 0.07585 in the KNN imputation data model with a Polynomial Kernel and a tolerance of 0.01.

Table 2. MSE and MAE Results of Tolerance Parameters

MSE Results Tolerance Parameter Analysis						
Imputation Data	**0.1**	**0.01**	**0.001**	**0.0001**	**0.00001**	**MIN**
ANN Linear	0.02190	0.02186	0.02177	0.02177	0.02178	0.02177
ANN RBF	0.02029	0.02017	0.02016	0.02016	0.02016	0.02016
ANN Poly	0.02184	0.02122	0.02119	0.02119	0.02119	0.02119
KNN Linear	0.01387	0.01561	0.01566	0.01566	0.01563	0.01387
KNN RBF	0.01973	0.07808	0.01507	0.01507	0.01565	0.01507
KNN Poly	0.01267	0.01279	0.01277	0.01277	0.01277	**0.01267**
Binning Linear	0.05840	0.05841	0.05829	0.05829	0.05829	0.05828
Binning RBF	0.22443	0.23898	0.24025	0.24025	0.24024	0.22443
Binning Poly	0.05901	0.05925	0.05921	0.05921	0.05921	0.05901
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.06477	0.06441	0.06443	0.06442	0.06441	0.06441
ANN RBF	0.09651	0.09271	0.09270	0.09270	0.09270	0.09270
ANN Poly	0.06981	0.06543	0.06511	0.06510	0.06511	0.06510
KNN Linear	0.08358	0.07795	0.07793	0.07790	0.07792	0.07790
KNN RBF	0.10068	0.17573	0.10068	0.08766	0.08957	0.08766
KNN Poly	0.07683	0.07585	0.07597	0.07599	0.07599	**0.07585**
Binning Linear	0.19651	0.19605	0.19577	0.19575	0.19576	0.19575
Binning RBF	0.33112	0.33651	0.33709	0.33713	0.33713	0.33112
Binning Poly	0.19512	0.19600	0.19592	0.19588	0.19589	0.19512

3.3 Variable C Parameters Analysis

The value of parameter C indicates how much SVR optimization avoids forecasting errors in each training session. A larger value of parameter C will result in a wider hyperplane edge distance, so if the parameter C is small, it will narrow the hyperplane edge distance. There are very significant changes in the value of parameter C in Binning, RBF, and Polynomial. The smallest MSE result is in the KNN method with a polynomial kernel at a parameter value of C1e + 5 with an MSE result of 0.0126. Meanwhile, the smallest

MAE result is the MAE in the ANN method with a polynomial kernel at a parameter value of C1 with an MAE result of 0.0640.

Table 3. MSE and MAE results of parameter C

MSE Results Variable C Parameter Analysis						
Imputation Data	1	1.00E+02	1.00E+03	1.00E+04	1.00E+05	**MIN**
ANN Linear	0.0219	0.0218	0.0218	0.0201	0.0189	0.0189
ANN RBF	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202
ANN Poly	0.0216	0.0213	0.0212	0.0212	0.0212	0.0212
KNN Linear	0.0162	0.0162	0.0156	0.0137	0.0230	0.0137
KNN RBF	0.0161	0.0197	0.0197	0.0197	0.0197	0.0161
KNN Poly	0.0133	0.0128	0.0128	0.0127	0.0126	**0.0126**
Binning Linear	0.0584	0.0583	0.0583	0.0583	0.0686	0.0583
Binning RBF	0.0822	0.1441	0.2403	0.3444	0.5171	0.0822
Binning Poly	0.0605	0.0587	0.0592	0.0600	0.0605	0.0587
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.0643	0.0644	0.0644	0.0670	0.0760	0.0643
ANN RBF	0.0927	0.0927	0.0927	0.0927	0.0927	0.0927
ANN Poly	0.0639	0.0646	0.0651	0.0653	0.0651	**0.0639**
KNN Linear	0.0768	0.0768	0.0779	0.0856	0.1207	0.0768
KNN RBF	0.0883	0.1007	0.1007	0.1007	0.1007	0.0883
KNN Poly	0.0756	0.0759	0.0760	0.0761	0.0790	0.0756
Binning Linear	0.1965	0.1959	0.1958	0.1963	0.2185	0.1958
Binning RBF	0.2314	0.2893	0.3371	0.3780	0.4163	0.2314
Binning Poly	0.2034	0.1955	0.1959	0.1970	0.1979	0.1955

3.4 Epsilon Parameter Analysis

Epsilon in ϵ-SVR is a very easy parameter to understand. It indicates how much error range is allowed per event training civil. So, the range is usually 0 to MAX_ALLOWABLE _ERROR. In this study, the effect of epsilon changes on the SVM RBF test results was tested. The epsilon ranges tested were from 0.1, 0.001, 0.0001, 0.00001. From the test results, the smaller the epsilon range tested, the lower the MSE and MAE levels. Table 4 shows that the smallest MSE result in the KNN method with a polynomial kernel at an epsilon parameter value of 0.01 with an MSE result of 0.0126. Table 4 shows that the smallest MAE result in the ANN method with a linear kernel is an epsilon parameter value of 0.0001 with an MAE result of 0.0064.

Table 4. MSE and MAE results of epsilon parameter

MSE Results Tolerance Parameter Analysis						
Imputation Data	**0.1**	**0.01**	**0.001**	**0.0001**	**0.00001**	**MIN**
ANN Linear	0.0201	0.0213	0.0218	0.0218	0.0218	0.0201
ANN RBF	0.0243	0.0202	0.0202	0.0202	0.0202	0.0202
ANN Poly	0.0202	0.0208	0.0212	0.0212	0.0212	0.0202
KNN Linear	0.0146	0.0138	0.0139	0.0139	0.0141	0.0138
KNN RBF	0.0135	0.0186	0.0197	0.0199	0.0199	0.0135
KNN Poly	0.0133	0.0126	0.0128	0.0128	0.0128	**0.0126**
Binning Linear	0.0581	0.0583	0.0583	0.0583	0.0583	0.0581
Binning RBF	0.1534	0.2313	0.2403	0.2411	0.2411	0.1534
Binning Poly	0.0587	0.0592	0.0592	0.0591	0.0591	0.0587
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.1063	0.0650	0.0644	0.0644	0.0644	**0.0644**
ANN RBF	0.1297	0.0950	0.0927	0.0925	0.0925	0.0925

ANN Poly	0.1069	0.0665	0.0651	0.0651	0.0651	0.0651
KNN Linear	0.0936	0.0843	0.0836	0.0839	0.0837	0.0836
KNN RBF	0.0835	0.0972	0.1007	0.1011	0.1011	0.0835
KNN Poly	0.0871	0.0768	0.0760	0.0759	0.0759	0.0759
Binning Linear	0.1968	0.1957	0.1958	0.1957	0.1957	0.1957
Binning RBF	0.2937	0.3325	0.3371	0.3376	0.3376	0.2937
Binning Poly	0.1957	0.1956	0.1959	0.1958	0.1958	0.1956

3.5 Maximum Iteration Parameter Analysis

This iteration test is a test to see the overall MSE and MAE system. The greater the number of iterations the better the accuracy obtained, this is for MSE and MAE, but there is instability in the accuracy level, namely at the number of iterations to 500 and 1000 which increases and decreases. This is because the best MSE and MAE values have not been found. The best MSE value is infinite iteration (-1), polynomial kernel, KNN preprocessing, the MSE value is 0.0128. Figure 11 shows the smallest MAE value obtained in the Linear kernel, infinite iteration (-1), ANN preprocessing, the MAE value is 0.0644.

Table 5. MSE and MAE results of maximum iteration parameters

MSE Results Tolerance Parameter Analysis						
Imputation Data	50	100	500	1000	-1	**MIN**
ANN Linear	1.0104	0.2035	1.1029	0.0670	0.0218	0.0218
ANN RBF	0.1308	0.1019	0.0379	0.0256	0.0202	0.0202
ANN Poly	0.3008	0.3906	0.3318	0.0686	0.0212	0.0212
KNN Linear	0.1772	0.5277	0.3608	1.1266	0.0156	0.0156
KNN RBF	0.0201	0.0171	0.0167	0.0166	0.0197	0.0166
KNN Poly	0.0984	0.3730	0.5140	0.0936	0.0128	**0.0128**
Binning Linear	2.6803	1.7716	0.5838	0.3230	0.0583	0.0583
Binning RBF	0.1588	0.0904	0.0955	0.1004	0.2403	0.0904
Binning Poly	4.3704	0.8324	2.0746	1.7691	0.0592	0.0592
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.7988	0.3574	0.8517	0.2034	0.0644	**0.0644**
ANN RBF	0.3486	0.3064	0.1783	0.1345	0.0927	0.0927
ANN Poly	0.4437	0.5031	0.4808	0.2054	0.0651	0.0651
KNN Linear	0.3438	0.5968	0.4847	0.8563	0.0779	0.0779
KNN RBF	0.1054	0.0938	0.0905	0.0920	0.1007	0.0905
KNN Poly	0.2554	0.5192	0.6004	0.2417	0.0760	0.0760
Binning Linear	1.3238	1.1276	0.6221	0.4692	0.1958	0.1958
Binning RBF	0.3073	0.2482	0.2525	0.2534	0.3371	0.2482
Binning Poly	1.7827	0.7657	1.1732	1.1508	0.1959	0.1959

The experimental results show that changes in preprocessing methods in the SVR method affect the MSE and MAE error values as follows:

a. Artificial Neural Network (ANN)

The process of improving the value using ANN regression uses binary sigmoid in the hidden layer and purelin in the output. The concept used is to regress each feature, thus requiring a long time. The results of the data improvement process using ANN showed more optimal results in MAE than those using KNN and binning. The smallest MAE obtained after the regression process using SVR was 0.0614 with an RBF kernel, gamma of 0.001, C of 1000, tolerance of 0.001, and infinite iterations. Meanwhile, the best MSE value after the SVR process was 0.0189 with a linear kernel of C of 100,000, infinite iterations, and tolerance of 0.001.

b. K-Nearest Neighbor (KNN)

Conclusions on the effect of KNN preprocessing on parameter changes. From the results of testing the parameter changes made, optimal parameter values were generated for use in the KNN-corrected data,

based on the MSE and MAE evaluations. The smallest MAE obtained after the regression process using SVR was 0.0729 with an RBF kernel, gamma 1, C of 1000, epsilon 0.0001, tolerance 0.001, and unlimited iterations. Meanwhile, the best MSE value after the SVR process was 0.0125 with a Poly kernel, C of 1000, tolerance 0.001, epsilon 0.01, and unlimited iterations.

c. Binning

The process of improving values using the binning method involves taking the middle value. The results of the data enhancement process with binning show that the results are not as good as those using KNN and ANN because the replacement of missing errors only considers the middle value of each feature. The smallest MAE obtained after the regression process using SVR was 0.0583 with a Poly kernel, C1000, tolerance 0.001, epsilon 0.0001, and unlimited iterations. Meanwhile, the best MSE value after the SVR process was 0.1957 with a Poly kernel, C1000, tolerance 0.001, epsilon 0.0001, and unlimited iterations.

The research results are evident from changes in parameters such as gamma, tolerance, C1000, epsilon, and maximum iterations. The best MSE value was obtained using the KNN method. KNN is considered the best method for handling MSE because the last five features have the same replacement value. However, the ANN method performed best in MAE testing. MAE is a measurement of error by incorporating accuracy values. ANN has the best value because the spread between predicted and actual values is close, thus, the ANN's value is higher than KNN's.

4. CONCLUSION

Rainfall data was obtained from the Sultan Syarif Kasim II Meteorological Station in Pekanbaru, Riau City, from January 1, 2000, to October 13, 2018. The data consists of 10 features. The data was processed without separation between data taken from the rainy season and the dry season. The results of data processing used several preprocessing methods and SVR method parameters. The best MSE results to complement the data used in this study used the KNN and Poly kernel preprocessing parameters, C 1000, tolerance 0.001, epsilon 0.01 and infinite iterations. And the best MAE results used ANN preprocessing with RBF kernel, gamma 0.001, C 1000, tolerance 0.001 and infinite iterations.

REFERENCES

[1] J. Liu, C. Li, J. Tian, F. Yu, and Y. Wang, “Self-correcting multi-model numerical rainfall ensemble forecasting method,” 2017.
[2] J. Diez-Sierra and M. del Jesus, “A rainfall analysis and forecasting tool,” Environ. Model. Softw., vol. 97, no. November 2017, pp. 243–258, Nov. 2017.
[3] J. Diez-Sierra and M. del Jesus, “Subdaily Rainfall Estimation through Daily Rainfall Downscaling Using Random Forests in Spain,” Water, vol. 11, no. 1, p. 125, Jan. 2019.
[4] R. Mohd, M. A. Butt, and M. Zaman Baba, “SALM-NARX: Self Adaptive LM- based NARX model for the prediction of rainfall,” in 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, 2018, pp. 580–585.
[5] C. Dewi and M. Muslikh, “Perbandingan Akurasi Backpropagation Neural Network dan ANFIS Untuk Memprediksi Cuaca,” J. Sci. Model. Comput., vol. 1, no. 1, pp. 7–13, 2013.
[6] N. Ulinnuha and Y. Farida, “Prediksi Cuaca Kota Surabaya Menggunakan Autoregressive Integrated Moving Average (Arima) Box Jenkins dan Kalman Filter,” J. Mat. “MANTIK,” vol. 4, no. 1, pp. 59–67, May 2018.
[7] S. Adhy, A. Prasetio, B. Noranita, and R. Saputra, “Usability Testing of Weather Monitoring on Android Application,” in 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018, pp. 1–6.
[8] C. Dewi, D. P. Kartikasari, and Y. T. Mursityo, “Prediksi Cuaca Pada Data Time Series Menggunakan Adaptive Neuro Fuzzy Inference System (Anfis),” J. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 1, pp. 18–24, 2014.
[9] R. F. Rahmat, F. R. Nasution, Seniman, M. F. Syahputra, and O. S. Sitompul, “Implementation of bayesian model averaging on the weather data forecasting applications utilizing open weather map,” in IOP Conference Series: Materials Science and Engineering, 2018, vol. 309, pp. 1–10.
[10] BMKG, “Data Online - Pusat Database - BMKG,” dataonline.bmkg.go.id. [Online]. Available: https://dataonline.bmkg.go.id/home. [Accessed: 11-Dec-2018].
[11] Makkulau Makkulau, Susanti Linuwih, Purhadi Purhadi, and Muhammad Mashuri, “Pendeteksian Outlier dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode Likelihood Displacement Statistic-Lagrange,” J. Tek. Ind., vol. 12, no. 2, pp. 95–100, 2010.
[12] A. Zimek and P. Filzmoser, “There and back again: Outlier detection between statistical reasoning and data mining algorithms,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 6, p. e1280, Nov. 2018.
[13] I. Ben-Gal, “Outlier Detection,” in Data Mining and Knowledge Discovery Handbook, New York: Springer-Verlag, 1993, pp. 131–146.
[14] W. J. Faithfull, J. J. Rodríguez, and L. I. Kuncheva, “Combining univariate approaches for ensemble change detection in multivariate data,” Inf. Fusion, vol. 45, no. January 2019, pp. 202–214, Jan. 2019.
[15] E. Vilenski, P. Bak, and J. D. Rosenblatt, “Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks,” Comput. Electron. Agric., vol. 162, no. April, pp. 412–421, Jul. 2019.