P-ISSN: 2685-8738 E-ISSN: 2685-872X
Variance: Journal of Statistics and Its Applications
Volume 6 Issue 2 | October 2024
10.30598/variancevol6iss2page183-194

FILLING THE PRECIPITATION GAPS: ACCURATE IMPUTATION WITH SUPPORT VECTOR REGRESSION IN NORTH SULAWESI

Angelin Cahyaning[1], Brina Miftahurrohmah[2]*, Grandys Frieska Prassida[3], Tikno[4]
[1,2,3,4]Universitas Internasional Semen Indonesia
Jl. Veteran, Kb. Dalem, Sidomoro, Kebomas, Gresik, 61122, East Java, Indonesia

Corresponding Author's Email: *brina.miftahurrohmah@uisi.ac.id*

Abstract: *Incomplete precipitation data poses major challenges in accurate precipitation predictions, hindering the effectiveness of water resource management and disaster risk mitigation efforts in North Sulawesi, Indonesia. This research aims to develop a precipitation prediction model using Support Vector Regression (SVR) to handle missing data. The precipitation data used comes from BMKG and ERA5 stations. The results show that using the RBF kernel with parameters* $C = 1000$, $\varepsilon = 0.1$, $\gamma = 100$ *produces the best predictions, except Dtatiun Meteorologi Naha with* $\gamma = 1000$. *The best model is shown in the model evaluation RMSE of 0.099, MAE of 0.099, and* R^2 *of 0.999. The ability of SVR to capture precipitation trends is shown in the model evaluation results. The best model obtained is used for the missing data imputation process.*

Keywords**:** *Data Imputation, Model Evaluation, North Sulawesi, Precipitation, Support Vector Regression.*

1. INTRODUCTION

Indonesia is an archipelago that offers a wide range of geographical and climatic diversity. This condition makes understanding precipitation very important, especially in the context of agriculture as the main livelihood of most of the population, which can affect agricultural productivity and crop yields [1]. In 2022, North Sulawesi province had 256 rainy days, making it Indonesia's most rain-intensive region [2]. However, it was found that daily precipitation data sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG) still needs to be made available. The large number of missing precipitation data values significantly worsens the prediction accuracy [3]. Therefore, this research is focused on North Sulawesi province to explore precipitation and gain further understanding of the potential for climate change in the region and to overcome the problem of missing data by imputation.

North Sulawesi Province is in the northern part of Sulawesi Island, bordering the Philippines to the north, and the provincial government center is Manado City. The geographical location of North Sulawesi is between 0°N 3°N and 123°BT-126°BT, which makes it one of the regions north of the equator [4]. The province is an agricultural region that relies on the agricultural sector as the main pillar of livelihood and development; this sector covers various subsectors, including food crop production, horticulture, fisheries, livestock, and forestry [5].

Machine Learning is a form of artificial intelligence that allows software applications to improve predictive accuracy without complex programming. The role of Machine Learning includes the ability to collect data using algorithms that take historical data as input to predict new output values [6]. Support Vector Regression (SVR) is a high-dimensional non-linear mapping machine learning method that can be adapted to a small number of data sets [7]. SVR has high flexibility and accuracy, which is useful in prediction and data imputation [8]. This method can handle non-linear relationships between variables by fluctuating and non-linear precipitation data characteristics. In addition, its ability to handle incomplete data makes SVR an appropriate choice in the context of this study. Thus, this study uses a quantitative approach, namely the Support Vector Regression (SVR) model, to impute precipitation data. It is hoped that this research will provide a deeper understanding of precipitation in North Sulawesi province and simultaneously be the first step in improving the quality of climate data in the North Sulawesi province area.

Submitted: 2024-06-25 *Accepted: 2024-10-30*

2. METHODOLOGY

2.1. Data sources

The secondary data used in this study are BMKG station data and ERAS data. BMKG station data is obtained from online publications by the Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) through the website https://dataonline.bmkg.go.id. This website is a data service platform for the public [9]. While the ERAS data used in this study is the result of global modeling from the Copernicus Climate Change Service (C3S) ERA5 earth observation service, the data can be downloaded on the page https://cds.climate.copernicus.eu [10]. With precipitation data (RR) from January 1, 2020, to December 31, 2023, the research location is North Sulawesi Province, with four observation stations with precipitation data.

Table 1. Observation Stations' Names

No.	Station Observation	Regency / City	Location	
			Latitude	Longitude
1	Stasiun Geofisika Manado	Manado City	1.44340	124.83890
2	Stasiun Meteorologi Maritim Bitung	Bitung City	1.44310	125.17970
3	Stasiun Meteorologi Naha	Sangihe Islands	3.68594	125.52881
4	Stasiun Klimatologi Sulawesi Utara	North Minahasa	1.54585	124.92330

2.2. Research variables

The research variables are divided into response variables and predictor variables. The response variable in this study is BMKG's precipitation data, which is the focus of observation during the research period from January 1, 2020, to December 31, 2023. Precipitation measurements were carried out at four stations spread across North Sulawesi Province. These stations act as observation points to record precipitation data, which will then become the response variable in the analysis. The predictor variable in this study is the output of ERA5, which provides global modeling data related to various climate parameters, including precipitation. ERA5 data is a predictor variable that helps to understand and forecast precipitation patterns in North Sulawesi Province.

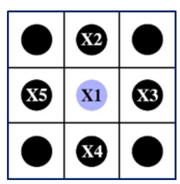

Figure 1. Determination of Grid Selection at Each Observation Location

The ERA5 variable used is the precipitation of the grids around the response variable point. The grid naming system can be seen in Figure 1. The blue dot (X1) is the grid at the observation location point.

2.3. Systematic Problem Solving

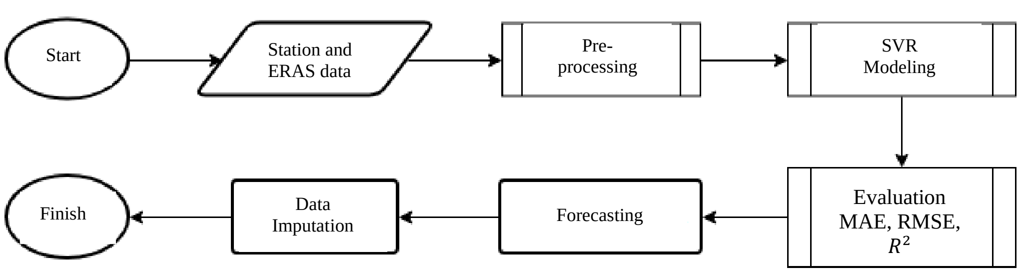

Figure 2. Research Stages

The analysis steps taken in this research include:

1. Data pre-processing

Data pre-processing was carried out to process ERA5 precipitation data from Copernicus and BMKG sources. This process includes extracting ERA5 data from Copernicus in netCDF format to CSV format and merging BMKG data from January 2020 to December 2023 with data from ERA5. BMKG data as response variable (Y) and ERA5 data as a predictor variable (X1, X2, X3, X4, X5). The next step is to separate the data, where the data is available for modeling, while the unavailable data is for use with the best model.

Table 2. Data Structures

Date	**Z**	**X1**	**X2**	**…**	**Xn**
January 1, 2020	………….	………….	………….	………….	………….
January 2, 2020	………….	………….	………….	………….	………….
⋮	………….	………….	………….	………….	………….
December 31, 2023	………….	………….	………….	………….	………….

The data structure was organized per observation station and variable. As part of the data preparation, ERA5 precipitation units were changed from m to mm to align with BMKG data.

2. Modeling using SVR

The second step is modeling using Support Vector Repression (SVR), which is performed using the Python programming language and libraries that support SVR implementation. The data is divided into training data (75.25%) and testing data (24.75%). Modeling started from January 1, 2020, to January 1, 2023, for turning data and continued with model testing using data from January 2, 2023, to December 31, 2023. The following is the formula for the non-linear SVR function used to predict precipitation values from the combined BMKG and ERAS datasets [11].

$$f\ (x) = \ w\ T\ \varphi(x) + \ b \tag{1}$$

Then, the equation for the structural minimization principle in SVR is used, aiming to find the weight vector w and bias & that produces a model with minimal prediction error [11].

$$min\frac{1}{2}\ \|w\|^2 + \ C\sum\ \sum_{i=1}^{n}(\varepsilon_k\ +\ \varepsilon_k^*) \tag{2}$$

The Radial Basis Function (RBF) kernel function is used to measure the similarity between 2 input vectors. This function produces higher values when the two input vectors are closer to each other in feature space [12].

$$(-\mathrm{z}\|\mathrm{x}\ -\ \mathrm{x}'\|^{\ 2}) \tag{3}$$

3. Model evaluation

Model evaluation is the third step, where model performance is measured using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) criteria. The evaluation results are used to select the best model, which is then used to make predictions.

1) Root Mean Squared Error (RMSE) [13]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}(\mathrm{x_t} - \hat{\mathrm{x}}_{\mathrm{t}})^2} \tag{4}$$

2) Mean Absolute Error (MAE) [14]:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}|\widehat{\mathrm{x}_{\mathrm{t}}} - \mathrm{x}| \tag{5}$$

3) Coefficient of Determination (R2) [15]:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(\hat{x}_t - x_t)^2}{\sum_{t=1}^{n}(\hat{x} - \bar{x})^2} \quad (6)$$

Information:

n = Number of samples or data

x = Actual value of x

$\hat{x}$ = Predicted value of x

4. Data Imputation

Support Vector Regression (SVR) is utilized in the data imputation process to fill in missing values and ensure dataset completeness. SVR was selected for its robust capability of predicting missing data points by identifying intricate patterns within the dataset. This methodological choice aims to preserve dataset integrity by minimizing information loss and facilitating accurate and reliable subsequent analyses. SVR's effectiveness in handling non-linear relationships between variables enables precise estimation of missing values, contributing to the overall validity and interpretability of research findings.

3. RESULTS AND DISCUSSION

3.1. Precipitation Modeling

Precipitation modeling is carried out by analysis at each observation station, which requires a special analysis approach to obtain accurate results. Researchers define 5 X variables in the ERA5 data with points located around the latitude and longitude of the observation station. The data was then analyzed using the Support Vector Regression (SVR) model with an RBF kernel consisting of variables (X1, X2, X3, X4, X5). The evaluation was carried out using three criteria, namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The best parameters will be selected based on the smallest value for RMSE, MAE, and the largest value for R^2.

3.1.1. Stasiun Geofisika Manado

Data collection at the Stasiun Geofisika Manado has a latitude point of 1.443 and a longitude point of 124.839. The calculation results on training and testing data can be seen in Table 3.

Table 3. Model Evaluation Calculation Results of Stasiun Geofisika Manado

Kernel	Parameter	Training Data Criteria			Testing Data Criteria		
		RMSE	MAE	R^2	RMSE	MAE	R^2
RBF	C = 1000, ε = 0.1, γ = 100	**0.099**	**0.099**	**0.999**	9.636	7.306	-0.013
	C = 10, ε = 0.9, γ = 1000	10.491	3.446	0.417	**9.538**	**6.353**	**0.007**
	C = 100, ε = 0.1, γ = 1000	1.039	0.133	0.994	9.640	7.423	-0.014

From the results of calculation results, the smallest error value in the RBF kernel precipitation data for the training data criteria is owned by the X1 – X5 model with parameters C = 1000, ε = 0.1, γ = 100, which produces an RMSE value of 0.099, MAE of 0.099, and the highest error value for the training data for R^2 is 0.999. The X1 owns the smallest error value in the testing data – X5 with parameters C = 10, ε = 0.9, γ = 1000, which produces an RMSE value of 9.538, MAE of 6.353, and the highest error value of testing data for R^2 is 0.007. Based on the model evaluation results in Table 3, it can be concluded that the best RBF Support Vector Regression (SVR) kernel parameters for precipitation prediction are parameters C = 1000, ε = 0.1, γ = 100.

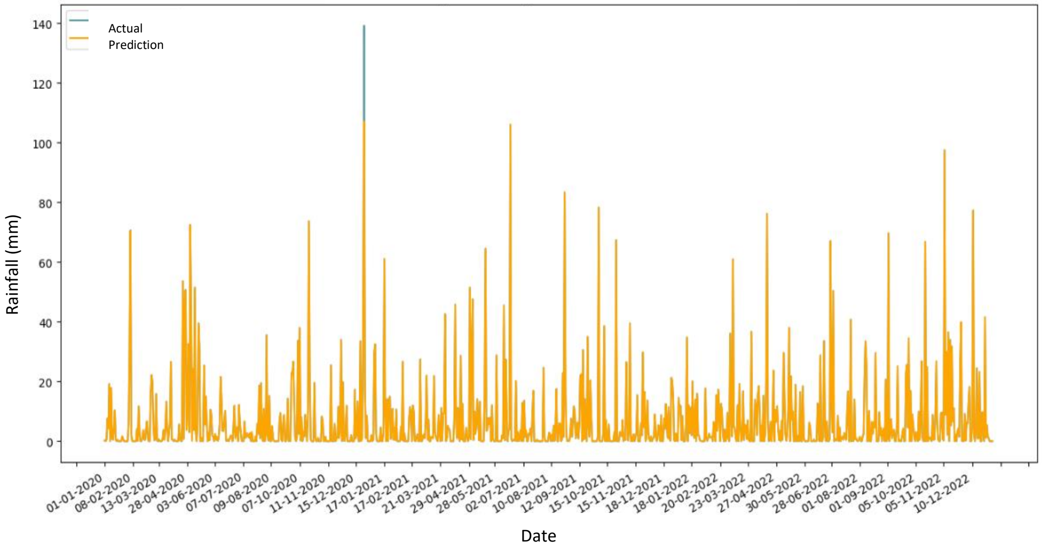

Figure 3. Training Data Graph of Stasiun Geofisika Manado

Figure 3 displays a training data graph consisting of actual precipitation data (blue line) from the Stasiun Geofisika Manado and prediction results (orange line) using the Support Vector Regression (SVR) model. The prediction lines tend to follow the actual data pattern, demonstrating the ability of the SVR model to capture precipitation trends. However, several points do not match. For example, on December 30, 2020, the actual precipitation data was 140 mm while the predicted data was 110 mm, showing the model variability level.

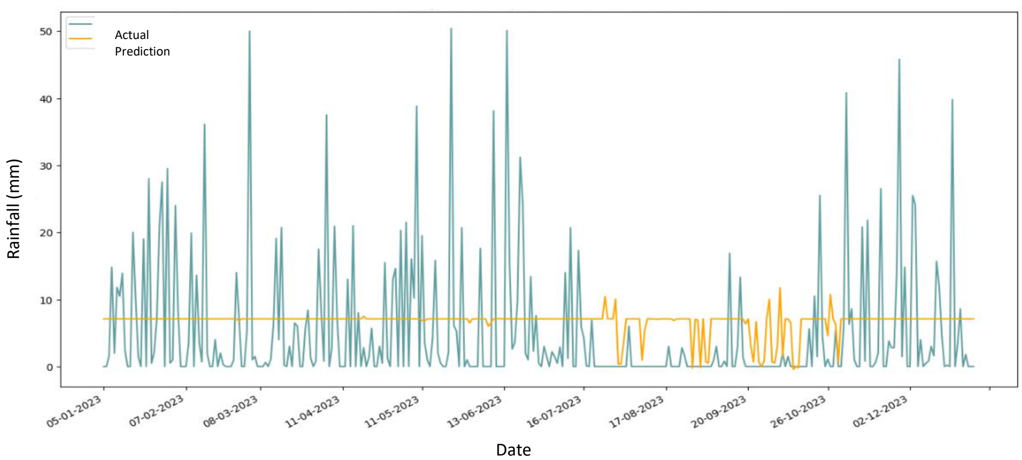

Figure 4. Testing Data Graph of Stasiun Geofisika Manado

Figure 4 displays a test data graph comparing the results of precipitation predictions (orange line) from the Stasiun Geofisika Manado using the SVR model with actual data (blue line). Although most prediction results are stable around value 7, fluctuations still occur, indicating a degree of uncertainty in the model predictions.

3.1.2. Stasiun Meteorologi Maritim Bitung

Data was collected at the Stasiun Meteorologi Maritim Bitung (Bitung City area), which has a latitude of 1.443 and a longitude of 125.179. Table 4 shows the calculation results for training and testing data.

Table 4. Model Evaluation Calculation Results of the Stasiun Meteorologi Maritim Bitung

Kernel	Parameter	Training Data Criteria			Testing Data Criteria		
		RMSE	MAE	R^2	RMSE	MAE	R^2
RBF	C = 10, ε = 0.1, γ = 100	11.350	2.826	0.383	**9.494**	**5.747**	**0.028**
	C = 1000, ε = 0.3, γ = 100	0.298	0.297	0.999	9.747	7.037	-0.024
	C = 1000, ε = 0.1, γ = 100	**0.099**	**0.099**	**0.999**	9.726	6.977	-0.019

The smallest error value in the training data results from parameters C = 1000, ε = 0.1, γ = 100 with an RMSE value of 0.099, MAE of 0.099, and the highest error value R^2 of 0.999. In contrast, the smallest error value in testing data results from the parameters C = 10, ε = 0.1, γ = 100 with an RMSE value of 9.494, MAE of 5.747, and the highest error value R^2 of 0.028. Based on the model evaluation results in Table 4, it can be concluded that the best RBF SVR kernel parameters for precipitation prediction are parameters C = 1000, ε = 0.1, γ = 100.

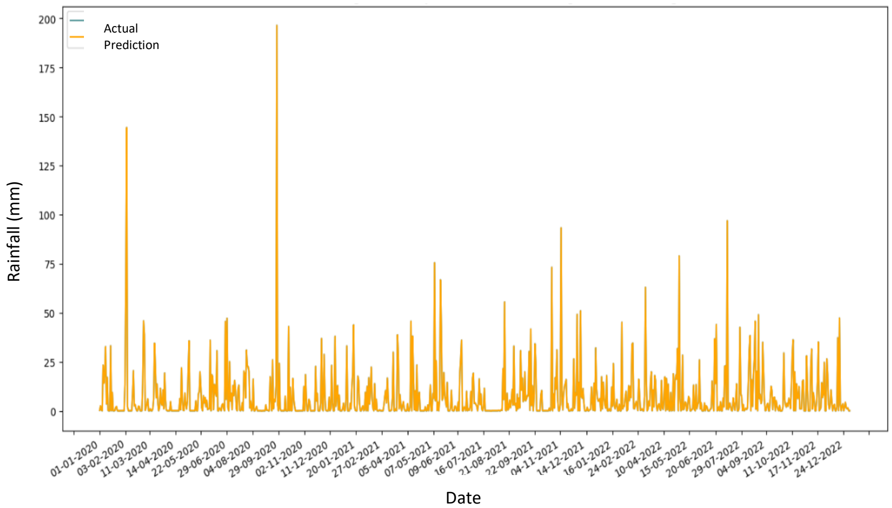

Figure 5. Training Data Graph of Stasiun Meteorologi Maritim Bitung

Figure 5 displays a training data graph consisting of actual precipitation data (blue line) and prediction results (orange line) using the SVR model. The prediction lines tend to follow the actual data pattern, demonstrating the ability of the SVR model to capture precipitation trends.

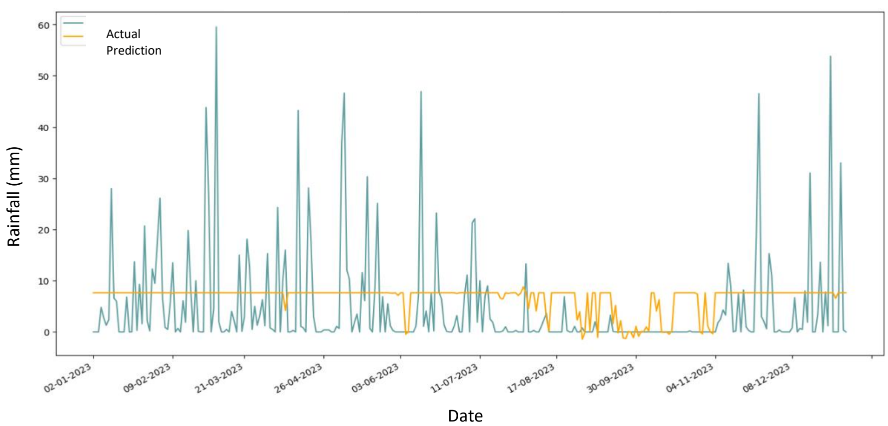

Figure 6. Testing Data Graph of Stasiun Meteorologi Maritim Bitung

Figure 6 displays a graph of testing data from predicted results from the SVR model, compared with actual data. Although most predicted results are stable around 7, fluctuations still occur.

3.1.3. Stasiun Meteorologi Naha

Data was collected at the Stasiun Meteorologi Naha (Sangihe Islands region), which has a latitude of 3.686 and a longitude of 125.529. Table 5 shows the calculation results for training and testing data.

Table 5. Model Evaluation Calculation Results of the Stasiun Meteorologi Naha

Kernel	Parameter	Training Data Criteria			Testing Data Criteria		
		RMSE	MAE	R^2	RMSE	MAE	R^2
RBF	C = 1000, ε = 0.1, γ = 1000	**0.099**	**0.099**	**0.999**	18.296	11.511	-0.018
	C = 10, ε = 10, γ = 0.1	17.717	11.265	0.169	18.118	11.759	0.002
	C = 10, ε = 0.1, γ = 10	16.141	5.096	0.311	**18.001**	**8.433**	**0.015**

The smallest error value in the training data results from parameters C = 1000, ε = 0.1, γ = 1000 with an RMSE value of 0.099, MAE of 0.099, and the highest error value R^2 is 0.999. Meanwhile, the smallest error value in the testing data results from the parameters C = 10, ε = 0.1, γ = 10 with an RMSE value of 18.001, MAE of 8.433, and the highest error value R^2 is 0.015. Based on the model evaluation results in Table 5, It is concluded that the best RBF SVR kernel parameters for precipitation prediction are parameters C = 1000, ε = 0.1, γ = 1000.

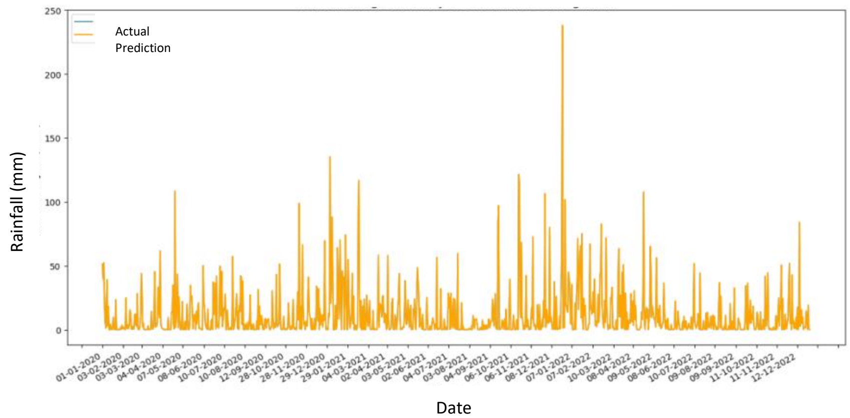

Figure 7. Training Data Graph of Stasiun Meteorologi Naha

Figure 7 displays a training data graph consisting of actual precipitation data (blue line) and prediction results (orange line) using the SVR model. Overall, the prediction lines follow the actual data pattern, demonstrating the ability of the SVR model to capture precipitation trends.

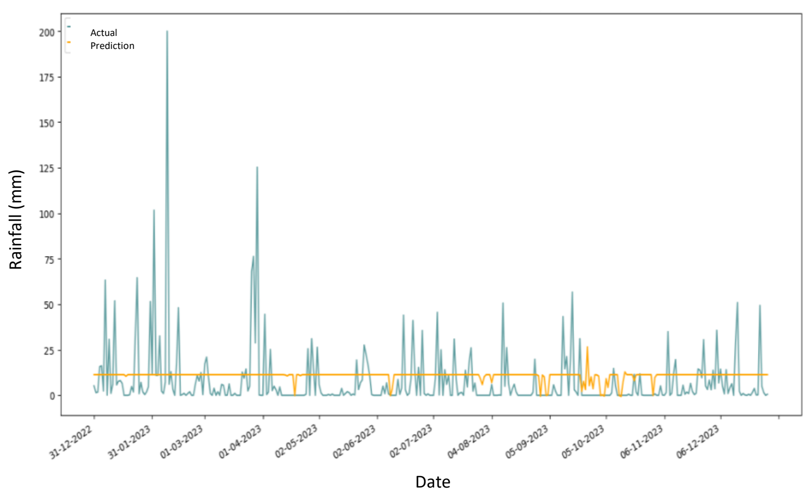

Figure 8. Testing Data Graph of Stasiun Meteorologi Naha

Figure 8 displays a graph of the predicted testing data (orange line) from the SVR model, compared with the actual data (blue line). Although most predicted results are stable around 10, fluctuations still occur.

3.1.4. Stasiun Klimatologi Sulawesi Utara

Data was collected at the Stasiun Klimatologi Sulawesi Utara (North Minahasa region), which has a latitude of 1.546 and a longitude of 124.923. Table 6 shows the calculation results for training and testing data.

Table 6. Model Evaluation Calculation Results of Stasiun Klimatologi Sulawesi Utara

Kernel	Parameter	Training Data Criteria			Testing Data Criteria		
		RMSE	MAE	R^2	RMSE	MAE	R^2
RBF	C = 100, ε = 0.9, γ = 10	2.626	1.010	0.983	20.838	11.873	0.032
	C = 1000, ε = 0.1, γ = 100	**0.099**	**0.099**	**0.999**	21.019	12.440	0.015
	C = 10, ε = 10, γ = 0.1	2.866	0.283	0.979	**20.804**	**12.644**	**0.035**

From the calculation results, the smallest error value obtained in the training data criteria is in the parameters C= 1000, ε = 0.1, γ = 100, which produces an RMSE value of 0.099, MAE of 0.099, and the highest error value of the training data for R^2 is 0.999. Meanwhile, the smallest error value in the testing data is the result of the parameters C = 10, ε = 10, γ = 0.1 with an RMSE value of 20.804, MAE of 12.644, and the highest error value in the testing data for $R^2$is 0.035. Based on the model evaluation results in Table 6, it can be concluded that the best RBF kernel parameters for precipitation prediction are parameters C = 1000, ε = 0.1, γ = 100.

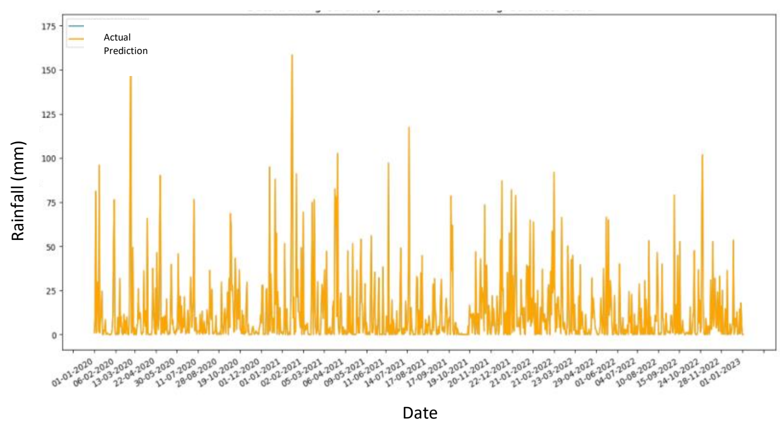

Figure 9. Training Data Graph of Stasiun Klimatologi Sulawesi Utara

Figure 9 displays a training data graph consisting of actual precipitation data (blue line) and prediction results (orange line) using the SVR model. Overall, the prediction lines tend to follow the actual data pattern, demonstrating the ability of the SVR model to capture precipitation trends. However, there are several points where the predictions do not match the actual data, such as on March 15, 2020, the actual data was 175 mm while the predicted data was 112 mm; this shows the level of variability or uncertainty in the model predictions.

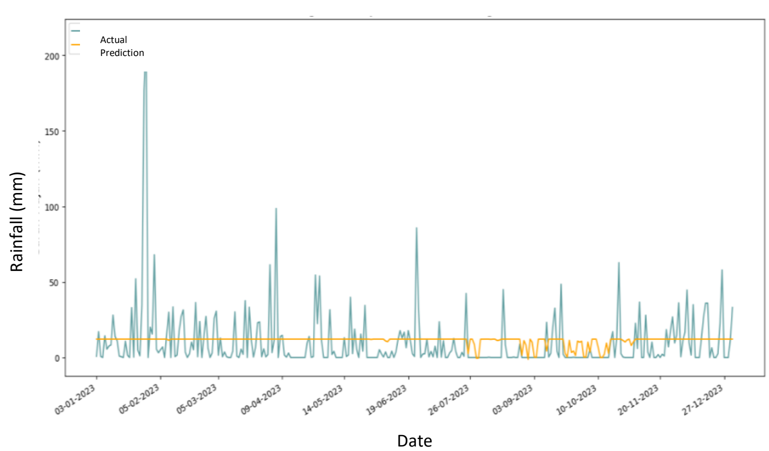

Figure 10. Testing Data Graph of Stasiun Klimatologi Sulawesi Utara

Figure 10 displays a graph of testing data from predicted results from the SVR model, compared with actual data. Although most predicted results are stable around value 12, fluctuations still occur.

3.2. Precipitation Prediction

Precipitation predictions are carried out using ERA5 data and the best parameter results from the RBF model at each station. Then, the ERA5 data is converted into mm (millimetres).

3.2.1. Stasiun Geofisika Manado

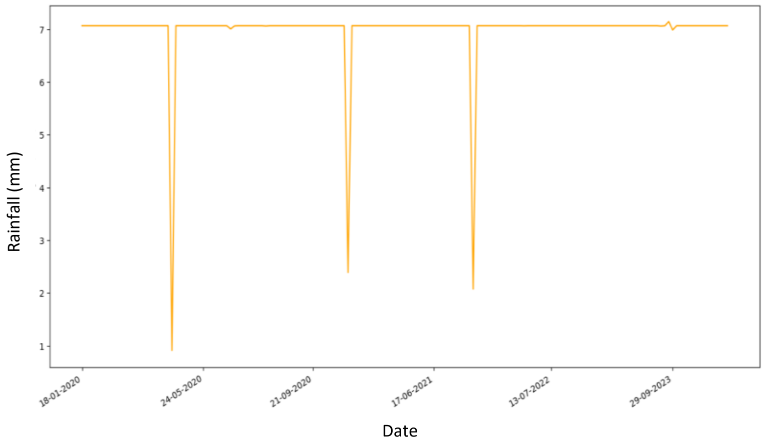

Figure 11. Unavailable Data Graph of Stasiun Geofisika Manado

The graph in Figure 11 displays the results of precipitation predictions from the Stasiun Geofisika Manado using the SVR model. Most predicted values are stable around 7.065, but some fluctuations occur, indicating variations in precipitation predictions. Nevertheless, the graph provides a general idea of the station's predicted precipitation pattern.

15	16-01-2020	0.0	3.965753	6.305968	4.337105	5.801023	3.433249
16	17-01-2020	0.0	6.658168	3.661198	5.388567	8.628433	4.807483
17	18-01-2020	NaN	0.475050	0.157416	0.422266	1.333595	0.484859
18	19-01-2020	0.0	0.759519	0.916934	0.777269	0.328845	0.966915
19	20-01-2020	0.0	1.404129	1.114989	2.850298	2.027252	0.410589
20	21-01-2020	1.6	1.235970	0.603037	1.728302	1.326589	0.325575
21	22-01-2020	0.7	4.057307	2.757810	5.314764	3.159524	1.332194
22	23-01-2020	0.0	0.905724	0.486260	1.377971	1.574156	0.181238
23	24-01-2020	NaN	0.806230	0.421332	0.747841	1.346208	0.215337
24	25-01-2020	0.0	0.111172	0.022888	0.114442	0.688051	0.050915
25	26-01-2020	0.0	0.227482	0.007941	0.266252	0.866487	0.047645
26	27-01-2020	NaN	0.616584	0.047645	0.553524	1.954383	0.178903

15	16-01-2020	0.0	3.965753	6.305968	4.337105	5.801023	3.433249
16	17-01-2020	0.0	6.658168	3.661198	5.388567	8.628433	4.807483
17	18-01-2020	7.1	0.475050	0.157416	0.422266	1.333595	0.484859
18	19-01-2020	0.0	0.759519	0.916934	0.777269	0.328845	0.966915
19	20-01-2020	0.0	1.404129	1.114989	2.850298	2.027252	0.410589
20	21-01-2020	1.6	1.235970	0.603037	1.728302	1.326589	0.325575
21	22-01-2020	0.7	4.057307	2.757810	5.314764	3.159524	1.332194
22	23-01-2020	0.0	0.905724	0.486260	1.377971	1.574156	0.181238
23	24-01-2020	7.1	0.806230	0.421332	0.747841	1.346208	0.215337
24	25-01-2020	0.0	0.111172	0.022888	0.114442	0.688051	0.050915
25	26-01-2020	0.0	0.227482	0.007941	0.266252	0.866487	0.047645
26	27-01-2020	7.1	0.616584	0.047645	0.553524	1.954383	0.178903

Figure 12. Data Imputation of Stasiun Geofisika Manado

Next, unavailable data is handled using the imputation method using predicted values from Support Vector Regression (SVR). As shown in Figure 12, data unavailable on January 18 2020, was replaced with a value of 7.1, which is the prediction result from the model.

3.2.2. Stasiun Meteorologi Maritim Bitung

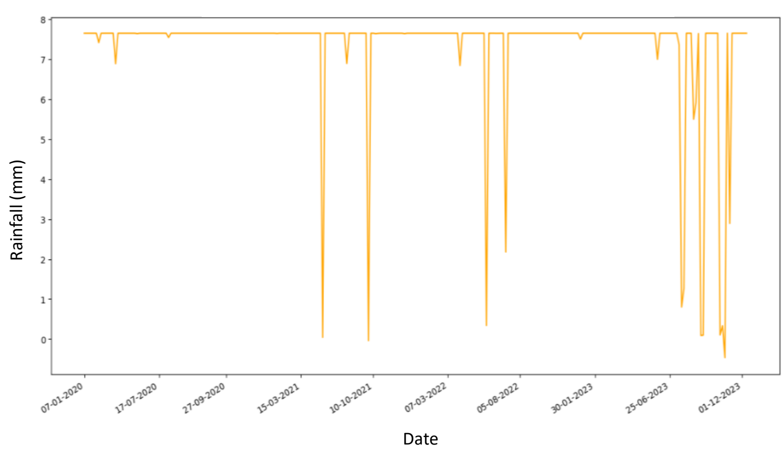

Figure 13. Unavailable Data Graph of Stasiun Meteorologi Maritim Bitung

The graph in Figure 13 displays the results of precipitation predictions from the Stasiun Meteorologi Maritim Bitung using the SVR model. Most predicted values are stable around 7.645, but some fluctuations occur, indicating prediction variations over time.

	Tanggal	RR (Z)	X1	X2	X3	X4	X5
0	01-01-2020	0.2	5.587088	9.576197	7.831078	5.169493	7.971211
1	02-01-2020	2.5	6.689932	6.934697	5.073268	4.359059	15.286135
2	03-01-2020	0.3	4.469764	4.573929	3.091793	5.723951	12.912287
3	04-01-2020	0.0	7.281758	11.294222	11.289084	7.814262	6.519904
4	05-01-2020	23.3	9.569657	11.508158	11.562343	9.321623	14.851723
5	06-01-2020	14.2	0.366680	1.299029	1.149555	0.044842	0.405917
6	07-01-2020	NaN	22.726709	34.267567	22.742124	16.771073	26.013754
7	08-01-2020	19.5	23.822080	20.625187	36.657292	28.208232	19.417243
8	09-01-2020	32.7	19.146789	22.616940	20.163216	14.984849	20.236554

	Tanggal	RR (Z)	X1	X2	X3	X4	X5
0	1/1/2020	0.2	5.587088	9.576197	7.831078	5.169493	7.971211
1	2/1/2020	2.5	6.689932	6.934697	5.073268	4.359059	15.286135
2	3/1/2020	0.3	4.469764	4.573929	3.091793	5.723951	12.912287
3	4/1/2020	0.0	7.281758	11.294222	11.289084	7.814262	6.519904
4	5/1/2020	23.3	9.569657	11.508158	11.562343	9.321623	14.851723
5	6/1/2020	14.2	0.366680	1.299029	1.149555	0.044842	0.405917
6	7/1/2020	7.6	22.726709	34.267567	22.742124	16.771073	26.013754
7	8/1/2020	19.5	23.822080	20.625187	36.657292	28.208232	19.417243
8	9/1/2020	32.7	19.146789	22.616940	20.163216	14.984849	20.236554

Figure 14. Data Imputation of Stasiun Meteorologi Maritim Bitung

Next, unavailable data is handled using the imputation method using predicted values from Support Vector Regression (SVR). As shown in Figure 14, data unavailable on January 7, 2020, was replaced with a value of 7.6, which is the prediction result from the model.

3.2.3. Stasiun Meteorologi Naha

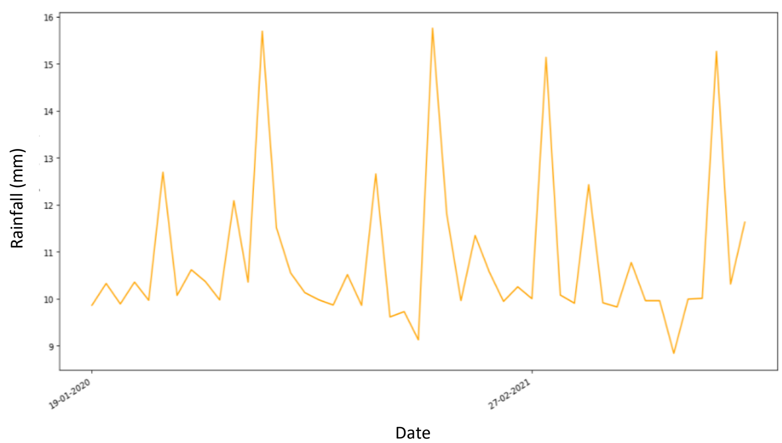

Figure 15. Unavailable Data Graph of Stasiun Meteorologi Naha

Using the prepared model, the graph in Figure 15 shows the precipitation prediction results (orange line) from the Stasiun Meteorologi Naha (Sangihe region). The prediction results are significantly variable, ranging from 9.124 to 15.693. The results indicate a considerable degree of fluctuation in precipitation predictions over time. Nevertheless, the graph provides a general idea of the predicted precipitation pattern from the station.

14	15-01-2020	1.0	0.300351	0.248969	0.569873	0.436747	0.162554
15	16-01-2020	0.0	7.363502	7.874986	6.282613	7.609202	7.619478
16	17-01-2020	9.5	4.312815	3.801331	3.282840	3.811607	5.270388
17	18-01-2020	2.8	0.001401	0.011211	0.001401	0.002803	0.004204
18	19-01-2020	NaN	1.131805	1.397590	0.942158	1.108916	1.071547
19	20-01-2020	0.0	1.238772	2.158977	0.800157	0.708137	2.051074
20	21-01-2020	0.0	0.732427	0.718413	0.747374	0.575011	0.414793
21	22-01-2020	NaN	2.175793	2.334143	2.100588	1.555005	1.514834
22	23-01-2020	23.5	2.476611	2.231846	2.662987	2.045936	1.649828
23	24-01-2020	0.0	1.495215	2.283695	1.285950	1.246246	1.903001

14	15-01-2020	1.0	0.300351	0.248969	0.569873	0.436747	0.162554
15	16-01-2020	0.0	7.363502	7.874986	6.282613	7.609202	7.619478
16	17-01-2020	9.5	4.312815	3.801331	3.282840	3.811607	5.270388
17	18-01-2020	2.8	0.001401	0.011211	0.001401	0.002803	0.004204
18	19-01-2020	9.7	1.131805	1.397590	0.942158	1.108916	1.071548
19	20-01-2020	0.0	1.238772	2.158977	0.800157	0.708137	2.051074
20	21-01-2020	0.0	0.732427	0.718413	0.747374	0.575011	0.414793
21	22-01-2020	10.3	2.175793	2.334143	2.100588	1.555005	1.514834
22	23-01-2020	23.5	2.476611	2.231846	2.662987	2.045936	1.649828
23	24-01-2020	0.0	1.495215	2.283695	1.285950	1.246246	1.903001

Figure 16. Data Imputation of Stasiun Meteorologi Naha

Next, unavailable data is handled using the imputation method using predicted values from Support Vector Regression (SVR). As shown in Figure 16, data unavailable on January 19, 2020 was replaced with a value of 9.7, which is the prediction result from the model.

3.2.4. Stasiun Klimatologi Sulawesi Utara

At the Stasiun Klimatologi Sulawesi Utara (North Minahasa region), precipitation is predicted using the Support Vector Regression (SVR) model with the RBF kernel. The results of precipitation predictions are shown in Figure 18, which shows fluctuations every year from 2020 to 2023. However, there will be a significant increase in 2022.

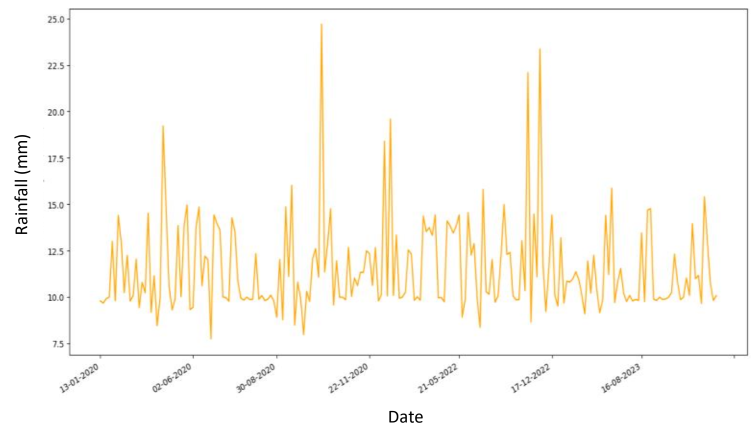

Figure 17. Unavailable Data Graph of Stasiun Klimatologi Sulawesi Utara

The graph in Figure 17 displays precipitation predictions (orange line) from the Stasiun Klimatologi Sulawesi Utara. The prediction results show significant fluctuations, with values varying from 7.751 to 24.704. The fluctuation indicates considerable variation in precipitation predictions over time.

14	15-01-2020	0.0	4.126439	1.404596	2.403275	6.186388	2.972213
15	16-01-2020	2.5	4.337105	6.153224	3.471552	4.502928	3.965753
16	17-01-2020	1.8	5.388567	2.967075	1.836205	6.035979	6.658168
17	18-01-2020	8.4	0.422266	0.102764	0.038303	0.843131	0.475050
18	19-01-2020	0.0	0.777269	0.910862	0.932816	0.395174	0.759519
19	20-01-2020	0.6	2.850298	2.018377	2.949792	2.219234	1.404129
20	21-01-2020	0.8	1.728302	0.911329	1.080423	1.554071	1.235970
21	22-01-2020	0.3	5.314764	3.622428	3.298255	2.171121	4.057307
22	23-01-2020	NaN	1.377971	0.742703	0.806230	0.702999	0.905724
23	24-01-2020	NaN	0.747841	0.277463	0.035500	0.462905	0.806230

14	15-01-2020	0.0	4.126439	1.404596	2.403275	6.186388	2.972213
15	16-01-2020	2.5	4.337105	6.153224	3.471552	4.502928	3.965753
16	17-01-2020	1.8	5.388567	2.967075	1.836205	6.035979	6.658168
17	18-01-2020	8.4	0.422266	0.102764	0.038303	0.843131	0.475050
18	19-01-2020	0.0	0.777269	0.910862	0.932816	0.395174	0.759519
19	20-01-2020	0.6	2.850298	2.018377	2.949792	2.219234	1.404129
20	21-01-2020	0.8	1.728302	0.911329	1.080423	1.554071	1.235970
21	22-01-2020	0.3	5.314764	3.622428	3.298255	2.171121	4.057307
22	23-01-2020	9.7	1.377971	0.742703	0.806230	0.702999	0.905724
23	24-01-2020	9.9	0.747841	0.277463	0.035500	0.462905	0.806230

Figure 18. Imputation of Stasiun Klimatologi Sulawesi Utara

Next, unavailable data is handled using the imputation method using predicted values from Support Vector Regression (SVR). As shown in Figure 18, data unavailable on January 23, 2020, was replaced with a value of 9.7, which is the prediction result from the model.

4. CONCLUSION

This research concludes that the Support Vector Regression (SVR) model can produce accurate precipitation predictions by filling in BMKG precipitation data for the North Sulawesi Province region, which is not available. The model evaluation shows that the best parameters for SVR using the RBF kernel are $C = 1000$, $\varepsilon = 0.1$, and $\gamma = 100$, except Stasiun Meteorologi Naha, Sangihe Islands show that $C = 1000$, $\varepsilon = 0.1$, and $\gamma = 1000$. Thus, SVR can be an effective tool in predicting precipitation in this region. As a suggestion, you can consider using other approaches, such as random forests, for further development.

REFERENCES

[1] E. Surmaini, E. Runtunuwu, and I. Las, “Agriculturals Effort to Anticipate Climate Change,” *J. Litbang Pertan.*, vol. 30, no. 98, pp. 1–7, 2017.

[2] BPS, "Statistik Indonesia 2023," *www.bps.go.id*, 2023. .

[3] T. Kim, W. Ko, and J. Kim, "Analysis and impact evaluation of missing data imputation in day-ahead PV generation forecasting," *Appl. Sci.*, vol. 9, no. 1, 2019, doi: 10.3390/app9010204.

[4] BPKRI, "Pemerintah Provinsi Sulawesi Utara," *BPK RI Perwakilan Provinsi SULAWESI UTARA*, 2020. https://sulut.bpk.go.id/pemerintah-provinsi-sulawesi-utara/.

[5] S. S. Pelengkahu, P. Kindangen, and E. N. Walewangko, "Analisis Pengaruh Sektor Pertanian Terhadap Pertumbuhan Ekonomi di Provinsi Sulawesi Utara," *J. Pembang. Ekon. dan Keuang. Drh.*, vol. 22, no. 2, pp. 46–66, 2021.

[6] L. Tucci and E. Burns, "What Is Machine Learning and Why Is It Important?," *SearchEnterpriseAI*, 2021. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML.

[7] S. Huang, L. Tian, J. Zhang, X. Chai, H. Wang, and H. Zhang, "Support Vector Regression Based on the Particle Swarm Optimization Algorithm for Tight Oil Recovery Prediction," *ACS Omega*, vol. 6, no. 47, pp. 32142–32150, 2021, doi: 10.1021/acsomega.1c04923.

[8] P. Pedamkar, "Support Vector Regression," *EDUCBA*, 2020. https://www.educba.com/support-vector-regression/.

[9] BMKG, "Tentang Aplikasi Data Online," *dataonline.bmkg.go.id*, 2015. https://dataonline.bmkg.go.id/.

[10] Copernicus, "About CDS," *cds.climate.copernicus.eu*, 2014. https://cds.climate.copernicus.eu/about-c3s.

[11] N. A. F. Sulaiman, S. M. Shaharudin, S. Ismail, N. H. Zainuddin, M. L. Tan, and Y. A. Jalil, "Predictive Modelling of Statistical Downscaling Based on Hybrid Machine Learning Model for Daily Rainfall in East-Coast Peninsular Malaysia," *Symmetry (Basel).*, vol. 14, no. 5, 2022, doi: 10.3390/sym14050927.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 16, no. 4. 2007.

[13] I. Fitriyaningsih *et al.*, "KEJADIAN BANJIR BERBASIS WEB DENGAN MACHINE LEARNING DI DELI SERDANG WEB-BASED APPLICATION DEVELOPMENT FOR PREDICTING RAINFALL , WATER DISCHARGE , AND FLOOD USING MACHINE LEARNING METHOD IN DELI SERDANG Fokus besar penelitian Machine Learning adalah bagai," *J. Penelit. Komun. dan Opini Publik*, vol. 22, pp. 150–160, 2018.

[14] D. Ardiansyah, "Perbandingan Model Prediksi Radiasi Matahari Berbasis Mesin Pembelajaran Pada Stasiun Meteorologi Fatmawati Soekarno Bengkulu," *Megasains*, vol. 14, no. 1, pp. 26–32, 2023, doi: 10.46824/megasains.v14i1.129.

[15] A. J. Alnaqbi, W. Zeiada, G. G. Al-Khateeb, K. Hamad, and S. Barakat, "Creating Rutting Prediction Models through Machine Learning Techniques Utilizing the Long-Term Pavement Performance Database," *Sustain.*, vol. 15, no. 18, 2023, doi: 10.3390/su151813653.