Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

Improving Multiclass Rainfall Prediction with Multilayer Perceptron and SMOTE: Addressing Class Imbalance Challenges

Nita Cahyani[1*], Wardiana Adinda Putri[2], Rahmat Irsyada[3]
[1]Universitas Padjadjaran, Sumedang, Indonesia, [2] Univertas Nahdlatul Ulama Sunan Giri, Indonesia [3]Politeknik Negeri Subang, Indonesia
[1]nita.cahyani@unpad.ac.id, [2]ardianaadinda0025@gmail.com, [3]irsyada.rahmat@gmail.com

open access

***Corresponding Author**
Article History:
Submitted: 02-12-2024
Accepted: 18-12-2024
Published: 20-12-2024

Keywords:
Rainfall classification; Class imbalance; Multilayer Perceptron; SMOTE; Neural Network

ABSTRACT

Rainfall is a crucial climatic element that significantly influences weather patterns and human activities, particularly in agriculture and daily life. Accurately classifying rainfall is essential for predicting future precipitation levels and managing its impacts effectively. This study applies the Multilayer Perceptron (MLP), a neural network algorithm, to classify rainfall patterns using a dataset obtained from the BMKG website. The dataset exhibited a class imbalance issue, which necessitated the application of the Synthetic Minority Over-sampling Technique (SMOTE) to address this challenge. The research compared the performance of the MLP model with and without the use of SMOTE. Without SMOTE, the MLP achieved an accuracy of 75%, a sensitivity of 40.34%, a specificity of 86.15%, and an AUC of 63.25%. With SMOTE, the MLP showed improved balance, achieving an accuracy of 71.27%, a sensitivity of 71.14%, a specificity of 90.30%, and an AUC of 80.72%. Although the overall accuracy decreased slightly, the significant improvement in sensitivity and AUC highlights the effectiveness of SMOTE in addressing class imbalance. The results demonstrate that SMOTE enhances the model's ability to identify minority classes, leading to more balanced and reliable predictions. This study underscores the importance of robust data preprocessing techniques in developing effective predictive models, particularly in climate-related applications where accuracy and balance are critical.

INTRODUCTION

One of the most crucial concerns is climate, especially in Indonesia where the country has a tropical environment Climate directly affects human survival in a number of domains, such as agriculture, industry, health, and others. Climate is an idea defined as the culmination of daily weather patterns and atmospheric features at a specific location over a long period of time (Gunadi et al., 2022).

Rainfall is one of the components of climate Weather is also affected by rainfall. Rainfall is the amount of water that falls on the surface of flat land within a certain period of time, such as days, weeks, months, or years, and is measured in millimeters (mm) above the horizontal surface(Triangga, 2020).

Although rainfall cannot be determined with absolute precision, we can predict or estimate it. To provide accurate predictions, climatologists have to try and work hard to solve problems that are a big challenge (Pradipta, 2020) Rainfall itself affects human activities, especially in farming and daily life Therefore, it is very important to classify rainfall so that we can predict the amount of rainfall that will come, especially if we predict rainfall every day (known as daily rainfall) (Gede et al., 2022).

Algorithms for machine learning can be used in classification analysis. Large volumes of data are typically analyzed, found, and extracted using machine learning techniques in order to identify patterns and knowledge from the data (Kamal & Ramdhani, 2023). Data mining is the process of extracting valuable information from complicated acquired and processed data (Wiranata et al., 2023). The Artificial Neural Network (ANN) is one of the various techniques that can be utilized for classification.

Over time, in real data there are many situations where the number of instances in one class is much less than the number of instances in another class. This situation is commonly known as the imbalance class problem (Sutoyo & Fadlurrahman, 2020).

In the preprocessing stage of rainfall data in Tuban Regency, it is known that the research data set has an unbalanced class problem, where of the four categories used, namely cloudy, light, medrum and heavy, the heavy category has a much lower value than the other three categories Thus, a method is needed to overcome the problem. Synthetic Minority Oversampling Technique (SMOTE) is one of the most commonly used dataset resample methods in research to solve the imbalance class problem (Sutoyo & Fadlurrahman, 2020).

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

LITERATURE REVIEW

The research to be conducted is supported by an analysis of previous research findings. One way to use it is to compare previous research, both in terms of advantages and disadvantages, and to support the claims made in the research. Some studies that apply the SMOTE technique include Naomi Nessyana Debataraja (2020) combining SMOTE with the Classification And Regression Tree (CART) method for the analysis of the National Socio-Economic Survey (SUSENAS) showing the results of unbalanced data research by applying the SMOTE technique as a whole improves the results of model evaluation where CART without SMOTE has an accuracy value of 91.92%, sensitivity 36.36%, specificity 98.94% and AUC 88.14%, while CART with SMOTE has an accuracy value of 91.92%. SMOTE has an accuracy value of 80.86%, sensitivity of 67.05%, specificity of 94.31% and AUC of 94.35%.

Further research conducted by Nugroho & Rilvani (2023) on the case of company bankruptcy, combining SMOTE with the Random Forest Classifier method which shows the results of testing the SMOTE oversampling technique in the Random Forest Classifier algorithm improves classification performance by 7.40% where the accuracy value of the Random Forest Classifier algorithm without SMOTE is 88.30% and the Random Forest Classifier algorithm with SMOTE is 95.70%.

The above description's results indicate that SMOTE can address the issue of class imbalance. Accordingly, this study will examine how well Synthetic Minority Oversampling Technique (SMOTE) performs in addressing the issue of class imbalance in Tuban Regency's daily rainfall classification from 2019 to 2023. "Analysis of Multiclass Imbalance in Rainfall Classification with the Smote Neural network Method" was thus chosen as the research title.

METHOD

Finding a solution to the problem described earlier requires a structured framework to describe the stages of the flow that will be used to solve the problem. This structured framework has several stages, namely, data cleansing, data input, pre-processing and multilayer perceptron classification as seen in the image below.

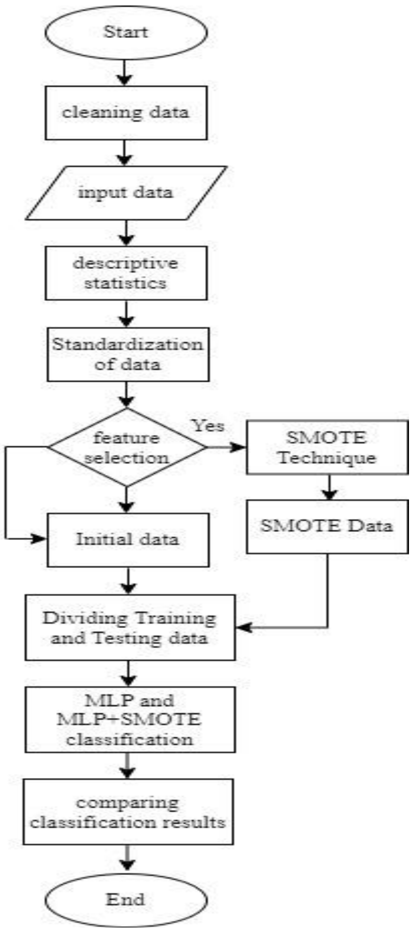

Figure 1. Flowchart

Cleaning Data

At this stage, rainfall data cleansing in Tuban Regency from 2019 to 2023 is carried out by detecting errors, deleting unmeasured data and improving the data structure.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

Input data

The next stage, enter the data that has been cleaned into the WEKA software for further processing.

Deskriptive statistic

At this stage, a characteristic analysis was carried out based on variables that are suspected to affect rainfall with descriptive statistics by looking for mean, maximum, minimum, and standard deviation.

Standartized of data

At this stage, standardise to bring different variables to the same scale with the range [0.1]. The data standardisation process is carried out by transforming the data into the form of Z-Score. Z-Score transforms or modifies data by creating new value ranges based on previously existing value ranges in the dataset. Z-score generates a new value based on the difference between the mean value and the standard deviation.

Feature selection

Feature selection is used to remove attributes or variables that are less relevant. Information Gain Feature Selection is one of the most popular selection feature techniques. This method uses the output variable to calculate the entropy of each attribute. The output value ranges from 0 to 1. Attributes that have a high information gain value and can be selected, while attributes with a lower information gain value and can be discarded.

SMOTE technique

At this stage, it aims to transform unbalanced data into balanced using the SMOTE technique. The SMOTE technique uses WEKA software by conducting a synthesis process to increase minority data by 115% with nearest neighbours which is 5.

Klasifikasi Multilayer Perceptron

At this stage, the data is divided into training data and data testing as well as K-Fold cross validation. By carrying out the multilayer perceptron classification process using the initial data and the multilayer perceptron classification process using the data that has been carried out by the SMOTE process. Then compare the classification results using the initial data with the classification using SMOTE data.

RESULT

This research uses the Multilayer perceptron method without SMOTE and Multilayer.perceptron with SMOTE using 80:20 data division. The learning rate used is 0.1 with a momentum value of 0.2 and uses 1 hidden layer with 9 neurons. Activation Function that Before classification will be done first standardized data and feature selection to remove variables or attributes less relevant The results can be seen as follows:

Data standardisation is used if the data has variable units that vary or differ and the data in this study has different units so it is necessary to standardise. The variables used in the study were minimum temperature, maximum temperature, average temperature, average humidity, duration of solar irradiation, maximum wind speed, wind direction at maximum speed, average wind speed and rainfall. All of the data is then standardised using WEKA software. The results of all data that have been standardised have been presented in Figure 2.

Viewer

Relation: Data Curah Hujan Tuban x-weka.filters.unsupervised.attribute.Standardize

No.	1: Tn Numeric	2: Tx Numeric	3: Tavg Numeric	4: RH_avg Numeric	5: ss Numeric	6: ff_x Numeric	7: ddd_x Numeric	8: ff_avg Numeric	9: **RR** Nominal
1	1.2460717...	0.9185489...	1.268095...	0.72833489...	0.3404133...	-0.663647...	-1.231749...	-0.202779...	Berawan
2	-0.337040...	-0.023694...	-0.296251...	0.72833489...	0.0219829...	-0.204261...	-1.158480...	-0.202779...	Sedang
3	-1.568350...	-0.212143...	-0.198479...	0.29916914...	-1.9947428...	-0.663647...	-1.158480...	-0.202779...	Ringan
4	0.5424662...	0.1647539...	0.094835...	0.44222439...	0.9065118...	1.173896...	-1.158480...	-0.202779...	Ringan
5	0.5424662...	0.7301002...	1.268095...	-1.5605491...	0.0573641...	-0.663647...	-1.158480...	-0.202779...	Berawan
6	0.5424662...	2.4261390...	1.952496...	-1.9897149...	0.5527003...	0.255124...	-1.158480...	1.0096149...	Berawan
7	0.5424662...	1.2954464...	1.170323...	-2.1327701...	1.2249422...	0.255124...	-1.158480...	1.0096149...	Berawan
8	1.4219731...	1.3896708...	2.050268...	-1.5605491...	1.2249422...	0.255124...	-1.158480...	-0.202779...	Berawan
9	0.5424662...	1.2954464...	1.952496...	-1.5605491...	1.4018480...	0.255124...	-1.158480...	1.0096149...	Berawan
10	1.4219731...	1.6723440...	2.050268...	-1.1313833...	1.2249422...	0.255124...	-1.158480...	1.0096149...	Berawan
11	1.4219731...	0.1647539...	0.877008...	0.15611388...	-0.6148778...	-0.663647...	-1.158480...	-0.202779...	Berawan
12	0.3665648...	0.3532026...	-0.198479...	0.87139015...	-2.3131732...	-0.204261...	-1.158480...	-1.415174...	Ringan
13	0.7183675...	-0.023694...	0.485921...	0.72833489...	-1.4994067...	-0.663647...	-1.158480...	-1.415174...	Berawan
14	0.8942689...	-0.212143...	0.388150...	0.72833489...	-0.2610662...	-0.204261...	-1.158480...	-1.415174...	Berawan
15	0.8942689...	0.0705295...	1.072551...	0.44222439...	0.6942249...	-0.204261...	-1.158480...	-1.415174...	Berawan
16	1.1581210...	0.5416514...	0.974780...	0.58527964...	0.5880814...	-0.204261...	-1.158480...	-0.202779...	Berawan
17	1.3340224...	0.5416514...	1.072551...	0.29916914...	-0.2964474...	-0.663647...	-1.158480...	-0.202779...	Berawan
18	-0.512942...	0.7301002...	0.192606...	0.44222439...	0.6942249...	-0.204261...	-1.158480...	-0.202779...	Ringan
19	-0.337040...	-0.212143...	0.388150...	0.29916914...	0.5880814...	-1.123034...	-1.158480...	-1.415174...	Sedang
20	1.4219731...	0.5416514...	1.170323...	-1.4174938...	0.2342699...	-1.123034...	-1.158480...	-0.202779...	Berawan
21	1.0701703...	2.9914853...	2.148040...	-1.7036043...	0.0573641...	-0.204261...	-1.158480...	-0.202779...	Berawan
22	0.5424662...	1.6723440...	1.659181...	-1.1313833...	1.1895611...	-0.204261...	-1.158480...	-0.202779...	Berawan
23	-0.337040...	-0.589041...	-0.198479...	0.72833489...	0.1281264...	0.255124...	-1.158480...	-0.202779...	Berawan
24	-1.216547...	-0.400592...	-0.980653...	1.15750065...	-1.7116936...	-0.204261...	-1.158480...	-0.202779...	Lebat

Add instance Undo OK Cancel

Figure 2. Standardized data

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

Information Gain Feature Selection is one of the popular feature selection techniques used. This technique determines the entropy of cach attribute using an output variable. The output value is between 0 and 1 Attributes with high gain values can be selected, but poor gain value attributes can be removed.

Table 1. Feature selection results

Variables	Description	InfoGain Value
X_4	Average humidity	0,2206
X_5	Length of sunshine	0,1438
X_1	Minimum temperature	0,1030
X_7	Wind direction at maximum speed	0,0978
X_2	Maximum temperature	0,0912
X_3	Average temperature	0,0671
X_8	Average wind speed	0,0174
X_6	Maximum wind speed	0,0158

In the test results in table 1, an arbitrary cutoff of 0.05 will be used as the limit for relevant attributes, if the InfoGain value 0.05 then the variable will be selected in this study, while if the InfoGain value < 0.05 then the variable will be discarded or eliminated. Based on this, the variables selected in this study are average humidity, length of sunshine, minimum temperature, wind direction at maximum speed, maximum temperature and average temperature While the variables average wind speed and is the maximum wind speed which has an InfoGain value <0.05 is not used or will be eliminated.

Multilayer Perceptron without SMOTE

In this test the data used for training data is 1349 data while for testing data 340 data Evaluation of the 80%: 20% data division model can be seen with the confusion matrix in Table 1.

Table 2. confusion matrix MLP classification

Actual Class	Predicted Class				Total
	Lightweight	Cloudy	Medium	Heavy	
Light	50	40	2	0	**92**
Cloudy	24	202	0	0	**226**
Medium	9	5	3	0	**17**
Heavy	5	0	0	0	**5**
Total	**88**	**247**	**5**	**0**	**340**

The results of the confusion matrix in Table 2 show that in the “light” category with a total sample of 92 data that is predicted correctly as much as 50, as well as 40 predicted as the “cloudy” category and 2 predicted as the “medium” category. In the “cloudy” category with a sample size of 226, 202 data were predicted correctly, and 24 data were predicted as the “light” category. In the “moderate” category with a sample size of 17, 3 data were predicted correctly, 9 data were predicted as the “light” category and 5 data were predicted as the “cloudy” category. In the “heavy” category with a sample size of 9 data, no data is predicted correctly, and there are 7 data predicted as the “light” category. Calculation of accuracy, sensitivity, specificity and AUC values, namely:

$$accuracy = \frac{\sum_{i=1}^{l} TP_i}{n} = \frac{255}{340} = 0{,}75$$

$$sensitivity = \frac{\sum_{i=1}^{g} \frac{TP_i}{TP_i + FN_i}}{g} = \frac{\frac{50}{92} + \frac{202}{226} + \frac{3}{17} + \frac{0}{5}}{4} = 0{,}4034$$

$$specificity = \frac{\sum_{i=1}^{g} \frac{TN_i}{TN_i + FP_i}}{g} = \frac{\frac{210}{248} + \frac{69}{114} + \frac{321}{323} + \frac{340}{340}}{4} = 0{,}8615$$

$$AUC = \frac{1}{2}(0{,}4034 + 0{,}8615) = 0{,}6325$$

Based on the above calculations, the accuracy value is 0.75 or 75%, the sensitivity value is 0.4034 or 40.34%, specificity value of 0.8615 or 86.15%, and AUC value of 0.6325. There are several categories based on the AUC classification value and the AUC value in this study of 0.6325 puts it in the poor classification group.

Teknik SMOTE

SMOTE technique uses WEKA software by conducting a synthesis process to increase minority data by 115% with nearest neighbours which is 5. For the "dense" minor class with a total of 25 data, it is necessary to replicate 5 times to balance the amount of data in the major class. For a "medium" minor class with a total of 92 data, replication is

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

required 3 times to balance the number of classes in the major class. As for the "light" minor class with a total of 452 data, it is necessary to replicate 1 time to balance the amount of data in the major class. A comparison of the results of the initial data with the data after the SMOTE can be seen in Figure 3.

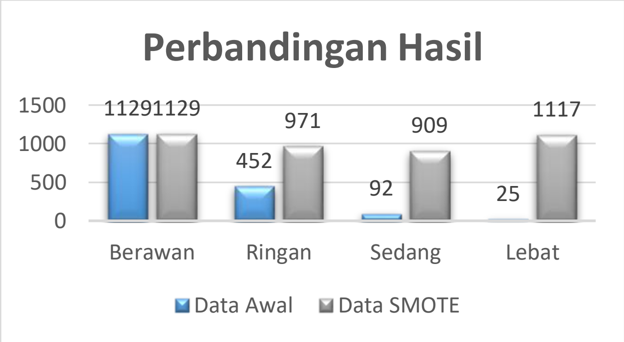

Figure 2. Data after SMOTE Technique

Based on the image above, the amount of data that was originally 1698 data will increase until the number of members in each class can be balanced. In addition to the dependent variable increasing, the number of data for each independent variable also increased along with the increase in the number of dependent variable data and in this study the number of SMOTE data was 4126 data.

Multilayer Perceptron with SMOTE

In this test, the data used for training data is 825 data while the rest is for testing data. Model evaluation can be seen with the confusion matrix in Table 3.

Table 3. Confusion matrix MLP+SMOTE classification

Actual Class	**Predicted Class**				**Total**
	Lightweight	**Cloudy**	**Medium**	**Heavy**	
Lightweight	94	53	45	16	**208**
Cloudy	25	186	17	2	**230**
Medium	21	10	130	26	**187**
Heavy	5	3	14	178	**200**
Total	**145**	**252**	**206**	**222**	**825**

The results of the confusion matrix in Table 4.20 show that in the “light” category with a total sample of 208 correctly predicted data as much as 94, as well as 53 predicted as the “cloudy” category, 45 predicted as the “medium” category and 16 predicted as the “heavy” category. In the “cloudy” category with a sample size of 230, 186 data were predicted correctly, and 25 data were predicted as the “light” category, 17 data were predicted as the “medium” category and 2 data were predicted as the “heavy” category. In the “moderate” category with a sample size of 187, 130 data were correctly predicted, and 121 data were predicted as the “light” category, 10 data were predicted as the “cloudy” category and 26 were predicted as the “heavy” category. In the “heavy” category with a sample size of 200, the correctly predicted data is 178, and there are 5 data predicted as the “light” category, 3 data predicted as the “cloudy” category and 14 data predicted as the “medium” category. Calculation of accuracy, sensitivity, specificity and AUC values, namely Translated with DeepL.com (free version):

$$accuracy = \frac{\sum_{i=1}^{l} TP_i}{n} = \frac{588}{825} = 0{,}7127$$

$$sensitivity = Sn = \frac{\sum_{i=1}^{g} \frac{TP_i}{TP_i + FN_i}}{g} = \frac{\frac{94}{208} + \frac{186}{230} + \frac{130}{187} + \frac{178}{200}}{4} = 0{,}7114$$

$$specifity = Sp = \frac{\sum_{i=1}^{g} \frac{TN_i}{TN_i + FP_i}}{g} = \frac{\frac{566}{617} + \frac{526}{595} + \frac{562}{638} + \frac{581}{625}}{4} = 0{,}9030$$

$$AUC = \frac{1}{2}(0{,}7114 + 0{,}9030) = 0{,}8072$$

Based on the above calculations, the accuracy value is 0.7127 or 71.27%, the sensitivity value is 0.7114 or 71.14%, specificity value of 0.9030 or 90.30%, and AUC value of 0.8072. There are several categories based on the AUC classification value and the AUC value in this study of 0.8072 puts it in the good classification group.

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

DISCUSSION

The results of the overall classification model evaluation comparison using 80%:20% testing with the Multilayer perceptron (MLP) method and the Multilayer perceptron (MLP) method with the Synthetic Minority Oversampling Technique (SMOTE) can be seen in Table 4.

Table 4. Evaluation of MLP and MLP+SMOTE models

Metode	**AUC**	***Sensitivity***	***Specificity***	***Accuracy***
MLP	63,25	40,34	86,15	75,00
MLP+SMOTE	80,72	71,14	90,30	71,27

Table 4 shows that when using the SMOTE handling technique to build the classification model, it results in a decrease in accuracy in each of the test scenarios. This is due to the fact that imbalanced data is usually classified as belonging to the majority class and results in high specificity values. Therefore, the accuracy value is less effective in measuring performance on imbalanced data. So this can cause a decrease in accuracy when using the SMOTE technique and shows the accuracy value is less effective for measuring classification performance on imbalance data. While the value of sensitivity, specificity and also the AUC value to experience an increase after using the SMOTE technique where the MLP method without Smote produces the AUC foor classfification category but when using the SMOTE technique becomes a good classification, so that the overall evaluation of the model shows that the SMOTE technique can be used to overcome the problem of multiclass imbalance in rainfall classification.

The performance of the accuracy value is less effective in using the model evaluation on the imbalance class, then the best test will be seen based on the AUC value, based on Table 4, it is known that the best classification test which shows the highest AUC value is obtained in the multilayer perceptron method with SMOTE, resulting in an AUC value of 80.72% so that it is included in the good classification category with a sensitivity value of 71.14% and a specificity value of 90.30%. The neural net model train from the best test will be presented in the form of visualization in Figure 3.

Based on Figure 3 above, it can be seen that the network has 6 inputs that represent each variable Then the number of hidden layers in the neural net is 1 with 9 neurons that have a weight on cach neuron and the final network which has 4 outputs representing each category also has a weight on cach output which will be presented in the form of table 5.

In Table 5, the weight that has the highest average value becomes the factor that most affects rainfall while the weight that has a negative value becomes a factor that has a low level of influence so that it can be ignored, based on this, it is known that the main variable that affects rainfall is variable namely maximum temperature with a value of 1.439. Then the variable that affects the second rainfall is X7, namely wind direction at maximum speed with a value of 0.302. Then the second variable that affects rainfall is, namely wind direction at maximum speed with a value of 0.302 The third variable that affects rainfall is which is the average humidity with a value of 0.060.

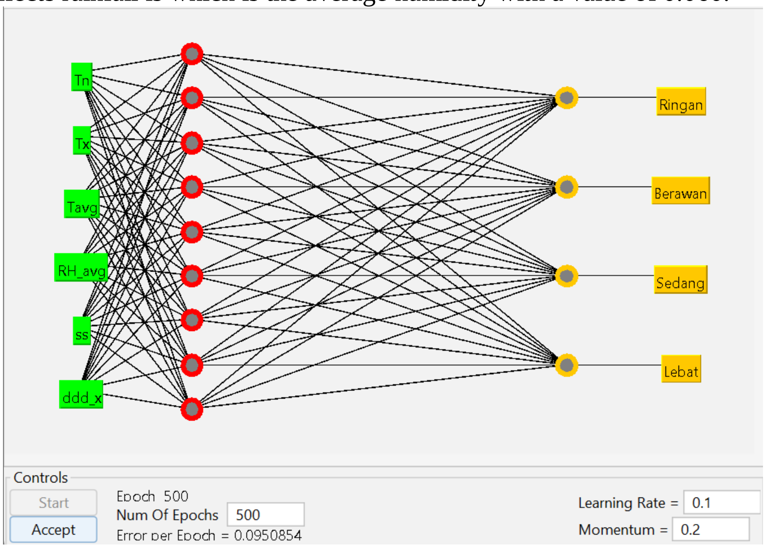

Figure 3. MLP model train

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

Table 5. weight value of each variable

Neuron	X_1	X_2	X_3	X_4	X_5	X_7
1	-0,365	-16,052	-3,180	9,041	0,406	4,549
2	-6,774	2,315	-0,379	-28,520	0,286	-5,019
3	6,623	16,664	-0,208	0,781	-13,295	-0,772
4	0,546	5,532	-16,581	-15,532	3,507	-1,538
5	-17,443	0,377	0,720	-0,495	-0,761	-1,450
6	9,628	3,324	1,630	24,715	-4,312	6,534
7	-10,639	-1,546	0,958	14,687	-0,884	0,523
8	14,307	0,603	2,679	-3,585	-0,277	-0,001
9	-4,769	1,729	-11,629	-0,555	-7,453	-0,109
Average	**-0,987**	**1,439**	**-2,888**	**0,060**	**-2,531**	**0,302**

CONCLUSION

Classification on imbalanced data using the Multilayer Perceptron (MLP) method obtained an accuracy value of 75% and an AUC value of 63.25% in the foor classification category, although the accuracy value is quite high but the classification is said to be poor based on the AUC value this is due to the fact that unbalanced data is usually classified as belonging to the majority class. So the accuracy value is less effective in measuring performance on imbalanced data. The Multilayer Perceptron (MLP)+SMOTE classification model with an AUC value of 80.72% falls into the good classification category Although the accuracy value has decreased when compared to Multilayer Perceptron (MLP) classification without SMOTE, but overall the model evaluation can be said that the SMOTE method is able to overcome imbalanced problems and improve classification performance. The main variable that affects rainfall is variable namely maximum temperature with a weight value of 1.439.

ACKNOWLEDGMENT

We would like to express our gratitude to all those who have supported and contributed to the writing of this article. The data used in this study is sourced from rainfall information obtained from the website of the Meteorology, Climatology, and Geophysical Agency (BMKG) and other related sources. We appreciate BMKG for providing valuable data that has greatly supported this research.

REFERENCES

Gede, I., Gunadi, A., Made, I., Oka Gunawan, A., Widastra, P. E., Candana, H., Ayu, I., Arnawa, W., Agoes Edo, K., & Putra, K. (2022). Klasifikasi Curah Hujan Harian Menggunakan Learning Vector Quantization. *Jurnal Ilmu Komputer Indonesia (JIK)*, *7*(2), 1–7.

Gumelar, G., Ain, Q., Marsuciati, R., Agustanti Bambang, S., Sunyoto, A., & Syukri Mustafa, M. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *SISFOTEK : Sistem Informasi Dan Teknologi*, 250–255.

Gunadi, I., Suseno, J. E., & Khuriati, A. (2022). Penentuan Curah Hujan Berdasarkan Input Cuaca Menggunakan Metode Logika Fuzzy Mamdani. *Jurnal Pengabdian*, *2(1)*, 155–159. https://doi.org/10.14710/Gading.

Imanwardhani, C. S. (2018). *Pendekatan Synthetic Minority Oversampling Technique dalam Menangani Klasifikasi Imbalanced Data Biner*. 13–15.

Kamal, N., & Ramdhani, Y. (2023). Optimasi Algoritma Neural Network berbasis Fitur Seleksi Menggunakan Algoritma Genetika Untuk Prediksi Curah Hujan. *E-PROSIDING TEKNIK INFORMATIKA*, *4*(2), 269–282.

Khaira, A. U. (2021). Analisis Sentiment Pada Tweet Dengan Tagar #Mahkamah Konstitusi Menggunakan Metode Multi Layer Perceptron. In *Repository Universitas Islam Riau*.

Mahmudah, K. R., Purnama, B., Indriani, F., & Satou, K. (2021). Machine learning algorithms for predicting chronic obstructive pulmonary disease from gene expression data with class imbalance. *BIOINFORMATICS 2021 - 12th International Conference on Bioinformatics Models, Methods and Algorithms; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, *February*, 148–153. https://doi.org/10.5220/0010316501480153

Naomi Nessyana Debataraja, R. D. P. S. W. R. (2020). Penerapan Synthetic Minority Oversampling Technique Dalam Mengatasi Data Tidak Seimbang Pada Metode Classification and Regression Tree. *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, *9*(1), 231–238. https://doi.org/10.26418/bbimst.v9i1.38949

Nugroho, A., & Rilvani, E. (2023). Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan. *Techno.Com*, *22*(1), 207–214. https://doi.org/10.33633/tc.v22i1.7527

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.5203

Nurdian, R. A., Mujib Ridwan, & Ahmad Yusuf. (2022). Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1), 24–32. https://doi.org/10.28932/jutisi.v8i1.4004

Pradipta, M. I. (2020). *Klasifikasi Curah Hujan Menggunakan Metode Ensemble Subset K-Nearest Neighbor*.

Ridho, I. I., Mahalisa, G., Sari, D. R., & Fikri, I. (2022). METODE NEURAL NETWORK UNTUK PENENTUAN AKURASI PREDIKSI HARGA RUMAH. *Technologia*, *13*(1), 56–58.

Ridhovan, A., & Suharso, A. (2022). Penerapan Metode Residual Network (Resnet) Dalam Klasifikasi Penyakit Pada Daun Gandum. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *7*(1), 58–65. https://doi.org/10.29100/jipi.v7i1.2410

Sutoyo, E., & Fadlurrahman, M. A. (2020). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *6*(3), 379. https://doi.org/10.26418/jp.v6i3.42896

Triangga, A. (2020). *Analisis Curah Hujan Dengan Pemodelan Deret Waktu Pada Das Walanae*. U N I V E R S I T A S B O S O W A.

Wiranata, A. D., Soleman, S., Irwansyah, I., Sudaryana, I. K., & Rizal, R. (2023). Klasifikasi Data Mining Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-Nn). *Infotech: Journal of Technology Information*, *9*(1), 95–100. https://doi.org/10.37365/jti.v9i1.164

Yustisio, D. (2022). *Perbandingan Metode extreme Learning Machine (ELM) dan Metode Multilayer Perceptron (MLP) Dalam Prediksi Jumlah Pasien Covid-19 Kota Semarang*.