



## Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)

Aditya Gumilar<sup>1</sup>, Sri Suryani Prasetyowati<sup>2</sup>, Yuliant Sibaroni<sup>3</sup>

<sup>1,2,3</sup>Informatics, School of Computing, Telkom University

<sup>1</sup>adityagumilar@student.telkomuniversity.ac.id, <sup>2</sup>srisuryani@telkomuniversity.ac.id, <sup>3</sup>yuliant@telkomuniversity.ac.id

### Abstract

*This study proposes several methods to analyze the performance of the hybrid machine learning method using Voting and Stacking on rainfall classification. The two hybrid methods will combine five classification methods, namely Logistic Regression, Support Vector Machine, Random Forest, Artificial Neural Network, and eXtreme Gradient Boosting. The data used is Bandung City rainfall data for the years 2005 until 2021. The hybrid method is classified as an ensemble, which means combining several individual classification models to improve the performance of the built model. Voting algorithm has weaknesses in imbalanced data, while stacking does not. The results show that by combining five machine learning methods on an imbalanced dataset, the Stacking algorithm obtains an accuracy value of 99.60%. Meanwhile, with the addition of the SMOTE technique, the accuracy increases to 99.71%. This is supported by the performance of the Stacking method which is superior because it takes the best classification value for each individual model and can overcome the imbalance. Model evaluation does not only focus on accuracy, but also precision, recall, and f1-score. The contribution of this research is to provide information about the best Hybrid method between Voting and Stacking in obtaining model performance results on rainfall classification.*

*Keywords: Rainfall, Machine Learning, Hybrid Methods, Classification, SMOTE.*

### 1. Introduction

Rainfall is the amount of water that falls due to cloud cover that has undergone a precipitation process and can be measured above a horizontal surface in units of millimeters (mm) in height [1]. BPS stated that the average amount of rainfall in the city of Bandung from 1980 to 2020 showed an increasing trend, although it fluctuated [2]. From 2005 to 2020, the amount of rainfall reached 3,038.31 mm with a surge in flood intensity occurring 220 times in the last 15 years [2][3]. In 2010 the amount of rainfall reached 322.5 mm which caused a surge in flood intensity to occur 25 times a year. In 2012, the amount of rainfall reached 209.3 mm and floods occurred 22 times. In 2016, the amount of rainfall reached 295.8 mm and caused the worst flooding in the city of Bandung in the last 20 years. In 2018-2020, rainfall reached 589.69 mm and caused the intensity of flooding to occur again as much as 142 times in three years.

Due to flooding due to high rainfall, people experience problems in conducting daily activities due to limited access to city roads, traffic jams, and damage to facilities and infrastructure. To reduce the impact and

damage, the community hopes for a system that can predict and analyze rainfall patterns based on historical data. One approach that can be used to analyze and predict is data mining [4]. However, the prediction results are not 100% accurate because each location, geography, and topography have different meteorological data. Several algorithms such as ANN, Random Forest, SVM, and Logistic Regression have investigated rainfall prediction. The performance of each of these algorithms varies greatly, so improving the performance of the model, it can be done by varying the number of samples used or combining different methods. Rainfall prediction continues to be a challenging task, therefore the selection of an appropriate method for classifying rainfall is very important in an area.

Rainfall classification has been carried out by several researchers [4], [5], [6], and [7] using individual machine learning methods. In research [4] applying the C5.0 algorithm using k-fold cross-validation and obtaining the highest accuracy of 92% on imbalanced data, while applying the smote technique the accuracy increased by 99%. Research [5] uses the PCA algorithm

for data processing, and SVM was used for classification. By using the parameter values  $C = 10000$  and  $\gamma = 0.5$ , the accuracy value is 65.28%. In research [6] using the Random Forest algorithm with two scenarios. The first scenario applies the 10-fold cross-validation technique and produces an accuracy value of 71.09%. In the second scenario, without cross-validation and obtaining an accuracy value of 99.45%. Another study applied the Logistic Regression algorithm using the undersampling technique, the highest result was the accuracy value reaching 84.24% after undersampling was carried out to overcome imbalanced data [7].

In studies [8], [9], and [10] still discuss the analysis of model performance on classification, but with different cases. In [8], applied the XGBoost algorithm to the classification of forest fires using feature importance. The accuracy results obtained are 89.52%. Research [11] apply the ANN algorithm in the medical field, especially in cancer research. While research [10] on the prediction of monthly weather classification. The results of the accuracy of the ANN method in research [10] using breast cancer data obtained is equal to 86.95% by applying 10 neurons to the input layer, sigmoid function, and 1 hidden layer. Research [10] compares the performance results of the ANN method based on the number of factors and the amount of data used. In this study, the ANN method is combined with backpropagation to calculate the weight of the ANN network and uses 11 hidden layers. The experimental results show that the prediction model using 5 factors and the amount of data in a 6-year period, obtains an accuracy rate of 83.33%. This value is higher than using 4 factors. In research [9], the hybrid method has not been applied to increase the accuracy value obtained. While in research [10], the number of factors used in building a predictive model is still too few.

Each method has advantages and disadvantages in building a classification model, depending on the type of data, the number of data samples used, and the number of attributes selected. Thus, to improve the performance of the built model, the hybrid method combines several individual classification models to build a high-performance model. In recent years, a combination of machine learning classification methods has been widely used in various studies, to produce accurate model performance in making predictions. However, not many studies have discussed the combination of classifications used in the case of rainfall. D. Sidik and T. sen in research [12] has proposed a stacking method on rainfall data by combining two machine learning methods, namely Naive Bayes and C4.5. The data used is daily climatological data for OPT Bandung, Bogor, Citeko, and Jatiwangi stations from 2000 to 2018, with 10 attributes. The test is carried out using data that has 5 classes, 2 classes, and 2 balanced classes. The results

show that the proposed stacking method can improve the performance of a model. The highest accuracy and f1-score values are found in the Majalengka dataset, which is 78.25% and 85.41% using 2 class targets. This research still has shortcomings because the accuracy value is still relatively low.

Research [13] and [14] proposed a combination stacking method used in the medical field. In [13] implemented a stacking technique by combining three machine learning methods, namely MLP, SVM, and LR. Several stages were carried out in this research, such as using the correlation technique for feature selection and implementing the AdaBoost method as a comparison of the performance of the proposed stacking method. The results obtained from this research are the performance of the stacking model can outperform the AdaBoost model and other individual methods with an accuracy of 78.2%, 1.66% higher than AdaBoost. This study shows that combining different methods can produce superior performance than combining a single method several times. The proposed stacking technique was also applied to different data sets, such as the diagnosis of heart disease and breast cancer, obtaining an accuracy of 80.2% and 97.4%, respectively. The stacking method performance metric measurement is still low when implemented in diabetes data, especially on the f1-score value of 59.4%.

A similar study was also conducted by S. Gupta and M. Gupta in [14] by proposing a stacking technique for the classification of a cervical cancer diagnosis. The dataset used has an imbalanced class because 95% is included in the healthy instance and 5% is cancer. To handle the imbalanced, the ROS technique is used to balance the number of classes in the data. Two feature selection approach techniques were also used to extract the most significant features. The proposed stacking method is applied with several machine learning approaches such as MLP, Gradient Boosting Classifier, Random Forest, and KNN. In addition, 2 other ensemble approaches such as majority voting and weighted voting are also used to measure the performance comparison generated by stacking. The results show that by using imbalanced data, the stacking architecture obtains an accuracy value of 95% which is better than other models. However, metric measurements such as precision, recall, and f1-score yield a value of 0%. To improve these results, several techniques are used to correct imbalanced data and two approaches to feature selection. By using balanced data or feature selection, measurement metrics such as accuracy, precision, recall, and f1-score are increased by more than 99%. The stacking technique proposed in this study is less than optimal when implemented on imbalanced data because of the significant difference in instances when it is implemented on balanced data.

The stacking method is also used in research [15] and [16] on unbalanced data. In research [15] proposed classifier ensemble stacking in predicting rock mass classification. 10 parameters are used because they have a stable phase after removing outliers. The proposed ensemble stacking method will be compared with seven individual classifications such as RF, GBDT, SVM, KNN, MLP, LR, and DT. By using the grid search method, the parameters for each classification can be optimized. The results show that the stacking method is effectively able to improve the performance of the classification method and has stronger learning abilities and generalizations for small samples and imbalanced data. In addition, it has a better performance than other individual classifications. With the influence of the SMOTE technique for minority class samples, the classification prediction model on the stacking algorithm produces an accuracy of 93%, 93% precision, 93% recall, 92.8 f1-score, and is better than the other seven methods. While [16] using the Stacking method and single classification methods such as K-NN, C4.5, SVM, and NN. The results show that the accuracy of the Stacking method by combining Tree C4.5 and SVM, obtains an accuracy value of 81% and is superior to the other four methods. Study [16] also explains that the Stacking method is a solution for datasets that have class imbalanced conditions.

In addition to stacking, other combination methods such as majority voting were also implemented in research [17] and [18]. In [17] discusses hybrid classification based on several classification methods to predict students who are entitled to the Smart Indonesia Program (PIP). This study focuses on the comparative analysis of hybrid voting methods with individual classification methods such as Artificial Neural Network, Naïve Bayes, K-nearest neighbor (KNN), and Iterative Dichotomiser 3 (ID3). To measure the proposed performance, this study uses four metric measurements. The results show that the hybrid classification using the voting method can produce an accuracy of 92%, while the best accuracy is produced by the ANN method with a value of 93%. Although the hybrid accuracy is lower than the individual classification method, the hybrid algorithm system is better and more consistent than other classifications with an F1-Score measurement of 94%. While in research [18], applied the hybrid voting method by combining three classification methods, namely Naïve Bayes, K-Nearest Neighbor, and Artificial Neural Network. The results show that the performance of the Voting method produces an accuracy value of up to 90% and is superior to other individual methods.

The hybrid stacking and voting method that has been proposed by several previous researchers are able to improve the performance of a classification model. Therefore, in research [19] combines stacking and ensemble voting methods by applying 15 classification

algorithms. The combination of the two hybrid methods is used in two datasets to predict faculty performance based on student responses. It was observed that in data 1 the proposed model has a performance of 75% and is better than other methods. In data 2 the proposed model has given a 2% higher performance. In this study, the performance of the stacking method combined with voting has a poor model performance, which is less than 80%.

Based on the research on the classification above, no research analyzes the comparison of the performance of the machine learning combination method used in imbalanced rainfall classification data. Thus, in this study, a comparative analysis of the two hybrid voting and stacking methods for rainfall classification will be carried out, combining five machine learning methods, and applying the SMOTE technique to overcome imbalanced data. The purpose of this study was to determine the best algorithm between voting and stacking in developing a rainfall classification model. Another goal that is expected is to find the best classification model that can be used by the community and government in estimating rainfall in the City of Bandung.

## 2. Research Methods

The system built is a process for classifying rainfall using two hybrid methods, with a combination of five machine learning methods. The hybrid model built has two scenarios, namely the use of data without the influence of the SMOTE technique and data with the SMOTE technique. The system design flow can be seen in Figure 1.

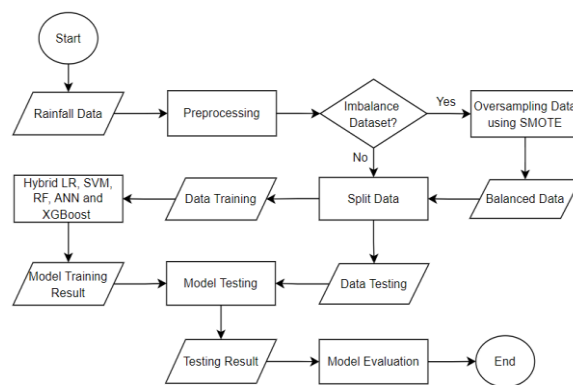


Figure 1. System Design Flow

### 2.1 Dataset

The data used is time series rainfall data and consists of 10 attributes, namely Date, Minimum Temperature (°C), Average Temperature (°C), Maximum Temperature (°C), Average Humidity (%), Rainfall (mm), Sunshine Length (hours), Maximum Wind Speed (m/s), Average Wind Speed (m/s), Wind Direction at Maximum Speed (°), and Most Wind Direction (°). The

data collected comes from the Meteorology Climatology Geophysics Agency (BMKG) Bandung City with daily details from 2005 until 2021. In this study, the date attribute will be divided into days, months, and years to form a new attribute. However, only the moon attribute will be added to the dataset because it has a strong correlation with the rainfall phenomenon.

## 2.2 Preprocessing

Data preprocessing is a technique in data mining that processes a set of raw data into data that produces information [7]. Preprocessing techniques are used to improve data quality, so that the results of the analysis obtained become more accurate, efficient, and can enter the stage of building a classification model.

## 2.3 Data Splitting

The data is divided into two partitions, with a ratio 80:20. Before entering the data mining stage, normalization will be conducted with the minmax method to change the shape of the data and accelerate the system learning stage. The concept of the minmax method is to perform a linear transformation on each feature to produce a decimal value with a range of 0 to 1 [20]. The minmax method formula can be calculated using equation (1).

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Where  $x$  is the original data [20].

## 2.4 Imbalanced Data and SMOTE method

Classification problems arise when the class being represented has an unbalanced number, this problem is known as an Imbalanced Dataset [4]. There are several ways to overcome unbalanced data, one of which is to apply the SMOTE technique [4]. SMOTE or also known as the Synthetic Minority Over-Sampling Technique is one of the oversampling approaches used to overcome imbalanced data types. SMOTE synthesizes minority classes through random data replication, so that the number of majority and minority classes in the data set is balanced [15].

## 2.5 Classification Process

Classification is a function to distinguish class objects based on data, with the aim that this function can be used to predict data that has no class or unknown class [21]. In classification, target data is a form of category.

### 2.5.1 Logistic Regression (LR)

Logistic Regression has a general form as a linear regression model and can be used to test the effect of factors that have numerical values on target variables with discrete outputs [22]. This method focuses on the relation of the independent variable  $(x_0, x_1, \dots, x_n)$  and

dependent variable 'Y' to predict data with discrete output values, such as 0 or 1 [7][23]. Multinomial Logistic Regression produces more than two outputs depending on the number of classes on the dependent variable. The following is the formula for Multinomial Logistic Regression [24]:

$$\Pr(Y = k) = \frac{e^{\mathbf{b}_k \cdot \mathbf{x}}}{1 + \sum_{k=1}^{K-1} e^{\mathbf{b}_k \cdot \mathbf{x}}} \quad (2)$$

Where  $Y$  is the result of a random variable,  $\mathbf{b}_k$  is the regression coefficient set related to class  $k$ , and  $\mathbf{x}$  is the observed climatic variable vector. Because the observation data is  $x$ , then the multinomial logistic regression produces a class label  $y$  as in equation (3):

$$y = \arg \max_k \Pr(Y = k) \quad (3)$$

### 2.5.2 Support Vector Machine

Support Vector Machine (SVM) algorithm is proven to be effective in overcoming regression and classification problems, because it can produce the best accuracy, efficiency, and solve overfitting problems [5][21]. The concept of the SVM method is to find a hyperplane that can divide two data sets from different classes. Suppose the training data consists of two classes  $[(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_i, y_i)]$  where  $x_i$  is the input vector, and  $y_i$  is the output. Then the formula for the binary classification problem is [5]:

$$f(x) = \text{sign}[w^T \cdot \phi(x) + b] \quad (4)$$

Where  $(x)$  is an input vector which is mapped into a non-linear feature space with function  $\phi(x)$ , while  $w$  and  $b$  are classification parameters. In representing the product results in the SVM method, you can use the kernel function as shown in equation (5):

$$K(x_i, y_i) = \begin{cases} x^T \cdot x' \rightarrow \text{linear} \\ (x^T \cdot x' + 1)^d \rightarrow \text{polynomial} \\ \exp(-\gamma \|x - x'\|^2) \rightarrow \text{RBF} \\ \tanh(\gamma x \cdot x' + C) \rightarrow \text{sigmoid} \end{cases} \quad (5)$$

If the data has a Lagrange multiplier that does not correspond to 0, then it is called a support vector. So, the classification formula is written as follows [5]:

$$f(x) = \text{sign} \left( \sum_{k=1}^m \alpha_k y_k K(x_k, x) + b \right) \quad (6)$$

Where  $m$  is the sum of the support vectors, while  $x_k$  represent the support vector. SVM algorithm can be formed when there are parameters  $C$  (misclassification tolerance),  $\gamma$  (gamma), Lagrange multiplier, and parameter  $b$  in equation (6).

### 2.5.3 Random Forest (RF)

Random Forest is a Bagging type classification method and is included in the supervised ensemble learning algorithm [6][7]. Random Forests have something in

common with Decision Trees. However, Random Forest has the principle that every tree is a weak learning, so to produce a strong learning model, random forest applies the concept of an ensemble tree. In Random Forest classification, Gini index is used as an attribute selection to measure the authenticity of the attribute relationship and its class. Gini index can be written as follows [25]:

$$\sum_{j \neq i} f(C_i, T) / |T| \cdot f(C_j, T) / |T| \quad (7)$$

where  $f(C_i, T) / |T|$  is the probability that the selected case belongs to the class  $C_i$ .

One of the advantages of the Random Forest method is that each tree in the training data will be built to reach the maximum depth, by combining features. Research [25] states that the performance of tree-based classification is influenced by the choice of pruning method and not based on the selection of attributes. The random forest classification consists of the number of  $N$ , where  $N$  is the number of trees that grow.

#### 2.5.4 Artificial Neural Network (ANN)

Artificial Neural Network is one of the popular machine learning techniques that simulates mathematical computational processes through biological neural networks [9]. The working principle of the ANN method is to create a system that can recognize patterns and adapt to new values in the data [18]. Feedforward Neural Network is the most common type commonly applied to ANN [9], because this type processes the input from the previous layer of neurons and sends the weight values as output to the next layer. In improving the performance of the ANN method, it is necessary to pay attention to the number of layers used, the number of neurons in the hidden layer and the relationship between each layer. The following is the formula used in the ANN method [26]:

$$y = g \left[ \sum_{j=1}^n w_{kj} f \left( \sum_{i=1}^n (w_{ji} x_i + b) \right) + b \right] \quad (8)$$

where  $x_1, x_2, x_3, \dots, x_i$  are input,  $y$  is the output,  $w_{j1}, w_{j2}, \dots, w_{jn}$  are synaptic weight in the hidden layer,  $w_{k1}, w_{k2}, \dots, w_{kn}$  are synaptic weight in the output layer, while bias or external threshold denoted by  $b$ . In the hidden layer there is an activation function which is denoted by  $f(x)$  and the activation in the output layer is denoted by  $g(x)$  [26].

In this study, the input neurons use 11 attributes which are denoted by  $X_1$  minimum temperature,  $X_2$  maximum temperature,  $X_3$  average temperature,  $X_4$  average humidity,  $X_5$  rainfall,  $X_6$  duration of sunlight,  $X_7$  maximum wind speed,  $X_8$  wind direction at maximum speed,  $X_9$  average wind speed,  $X_{10}$  the most wind

direction, dan  $X_{11}$  months when it rains. The number of hidden layers used is 23 based on the formula  $(2N + 1)$ , where  $N$  is the input variable [27].

#### 2.5.5 eXtreme Gradient Boosting (XGBoost)

XGBoost is one of the learning techniques that optimizes faster because it has been optimized with increasing gradients [8][28][29]. XGBoost has been widely used because of its fast, efficient, and scalable performance [29]. The principle of XGBoost is to achieve accurate prediction results through the iterative calculation of decision tree classification. Furthermore, XGBoost adds a regularization term to the cost function to reduce model variance and control model complexity to avoid overfitting. The cost function consists of a loss function ( $d$ ) and the regularization term ( $\beta$ ), so the formula for calculating the optimal value is written in equation (9)[30]:

$$\Omega(\theta) = \underbrace{\sum_{i=1}^n d(y_i, \hat{y}_i)}_{Loss} + \underbrace{\sum_{k=1}^K \beta(f_k)}_{Regularisasi} \quad (9)$$

Where  $n$  represents the number of object classes in the training sample,  $\hat{y}_i$  is predicted value,  $K$  is the number of trees to be produced, and  $f_k$  is number of trees from the ensemble. The regularization term is determined by equation (10).

$$\beta(f_t) = \gamma T + \frac{1}{2} \left[ \alpha \sum_{j=1}^T |c_j| + \lambda \sum_{j=1}^T c_j^2 \right] \quad (10)$$

Where  $\gamma$  is the minimum reduction of split loss,  $\lambda$  is the weight on the regularization term, and  $c$  is the weight relation for each leaf.

#### 2.5.5 Hybrid Classification

Hybrid is a combination method of two or more systems that run on the same function, taking into account the linear and nonlinear correlation structures [18]. Hybrid can improve model performance in classification, train many models to ensure errors have been made by one classification model, and predict output based on the highest probability of selecting a class as output [18].

##### 2.5.5.1 Voting Method

Voting is a type of meta-classification method that makes predictions by combining several individual classifiers. The Voting method involves the combination of several first-order predictive models to produce a second-level prediction model, where the results of the second level will outperform all of them [31]. The second level model takes advantage of the strengths of each model in the first level, where later each classification in Voting will be trained and tested with the same dataset in parallel. Voting classification

combines the predictions of N classifiers using hard Voting [31].

### 2.5.5.2 Stacking Method

Stacking or Stacked Generalization is one of the ensemble techniques where the output results from the first-level model set will be used as input to the second-level model [12]. Stacking differs from other ensemble techniques such as Bagging and Boosting. Because, Stacking can combine several classification models, whereas Bagging and Boosting only combine one model. Stacking approach can also overcome the imbalance in the data set used [16]. There are two stages of Stacking learning, namely each model will be trained using the same dataset to produce a base classifier for each model, this first stage is called a base learner or level 0 [12]. Base classifier or output obtained in stage one is used as input to create a new dataset in determining the predictions from the test data and providing the result. This second stage is known as meta-learners or level 1 [12].

### 2.6. Model Evaluation

To measure the performance of each method, this study evaluates the classification model using a confusion matrix. Parameters evaluated were accuracy, precision, recall, and f1-score. Accuracy is the ability of the model to predict the correct class, the accuracy formula is computable using an equation (11). Precision is the ratio of the class of rainfall that is predicted correctly, and the precision formulation can be calculated using equation (12). Recall is the ability of the model to predict the probability that a positive class will become positive, and the recall formula can be calculated using equation (13). F1-Score measures how much the system can predict the class correctly, the formula for F1-Score can be calculated by equation (14).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% \quad (11)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \quad (12)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (13)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (14)$$

### 3. Results and Discussions

This research was conducted using a rainfall dataset which has been divided into two partitions. Training data take 80% of the dataset to train each machine learning method, so that the classification probability is known in making decisions. While testing data is taken as much as 20% to test the algorithm model that has

been built. The classification model was developed with several scenarios.

### 3.1 Data Collection Scenarios

In this study, the dataset consists of 5 categories. The purpose of this category is to represent a class with a range of rainfall values that fall to a horizontal surface. Based on Table 1, the cloudy category has a total of 2086 classes, 1941 light rain classes, 422 moderate rain classes, 109 heavy rain classes, and 458 extreme rain classes.

Table 1. Number of Each Classes in Original Data

Range	Category	Class	Total
$Ch \geq 0$	Cloudy	0	2086
$0.5 \leq Ch \leq 20$	Light Rain	1	1941
$21 \leq Ch \leq 50$	Moderate Rain	2	422
$51 \leq Ch \leq 100$	Heavy Rain	3	109
$Ch > 100$	Extreme Rain	4	458

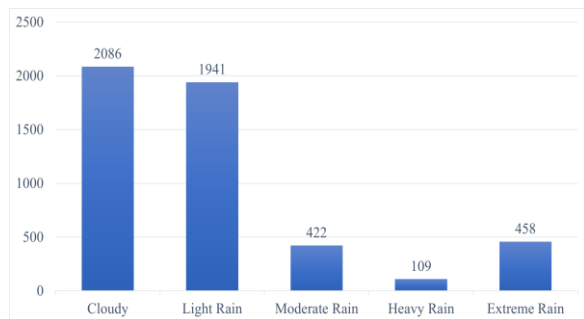


Figure 2. Total Number of Classes Without SMOTE Effect

Figure 2 shows that imbalanced data in the class can be observed. To avoid overfitting, the data will be processed using the SMOTE technique following section 2.4. The way it works is that the minority class will replicate a sample of the minority class randomly, so that the total for each class has the same number as in Table 2 and Figure 3.

Table 2. Number of Each Classes After Applying SMOTE

Range	Category	Class	Total
$Ch \geq 0$	Cloudy	0	2086
$0.5 \leq Ch \leq 20$	Light Rain	1	2086
$21 \leq Ch \leq 50$	Moderate Rain	2	2086
$51 \leq Ch \leq 100$	Heavy Rain	3	2086
$Ch > 100$	Extreme Rain	4	2086

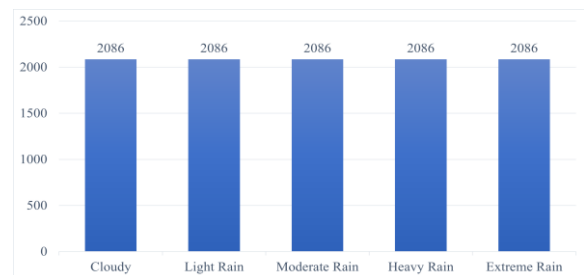


Figure 3. Total Number of Classes with SMOTE Effect



### 3.2 Testing Scenarios

The test scenario is divided into two parts, according to Table 3. The first is to implement five machine learning methods and two hybrid methods on imbalanced datasets. Furthermore, the dataset will be processed first to correct the imbalanced data and retested.

Table 3. Testing Scenarios

Testing	Data	Method	Hybrid
1	Data Without SMOTE effect	Logistic Regression, SVM, Random Forest, ANN, XGBoost	Voting, Stacking
2	Data with SMOTE effect	Logistic Regression, SVM, Random Forest, ANN, XGBoost	Voting, Stacking

### 3.2 Test Result

Not only accuracy parameters are used to measure the performance of each method, but other confusion matrix calculations such as precision, recall and f1-score are used together because there is a class imbalanced in the data.

Table 4. Model Performance without SMOTE Technique Effect

Method	Accuracy	Precision	Recall	F1-Score
LR	66.53%	88.01%	88.96%	88.14%
SVM	78.69%	79.52%	73.48%	75.97%
RF	84.96%	73.05%	52.50%	53.79%
ANN	86.65%	88.01%	88.96%	88.14%
XGBoost	81.67%	92.37%	48.64%	50.13%
Voting	82.37%	92.22%	61.05%	62.93%
Stacking	99.60%	97.34%	99.14%	98.14%

Table 5. Model Performance with SMOTE Technique Effect

Method	Accuracy	Precision	Recall	F1-Score
LR	72.87%	88.32%	88.39%	88.15%
SVM	90.89%	90.52%	90.65%	90.58%
RF	90.94%	90.83%	90.78%	90.58%
ANN	88.69%	88.32%	88.39%	88.15%
XGBoost	91.23%	91.21%	91.43%	90.95%
Voting	94.30%	94.17%	94.14%	94.14%
Stacking	99.71%	97.72%	99.71%	98.72%

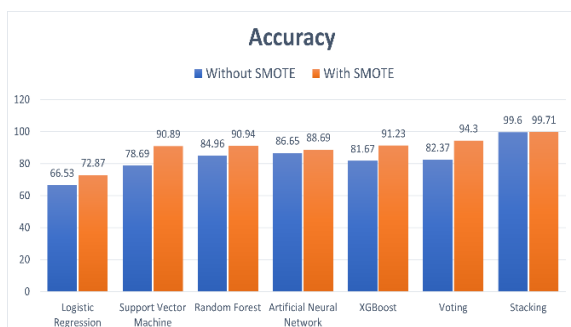


Figure 4. Evaluation of Model Accuracy Before and After Applying the SMOTE Technique

When training on imbalanced datasets, the model will be bias towards the majority class only. Thus, the performance of each method resulted in incorrect prediction models and less than optimal accuracy values as shown in Table 4. Voting method that should be able to improve model performance but has a lower accuracy value than other individual classification methods such as Random Forest and ANN. This is influenced by the prediction error of each model, and the existence of imbalanced datasets which causes the Voting method to be less effective in building a classification model. While Stacking method has advantages in handling imbalanced data as described in research [16]. Thus, the performance of Stacking method has a higher level of accuracy than other methods. After using SMOTE, the data became balanced, and the accuracy of each model increased as shown in Table 5 and Figure 4.

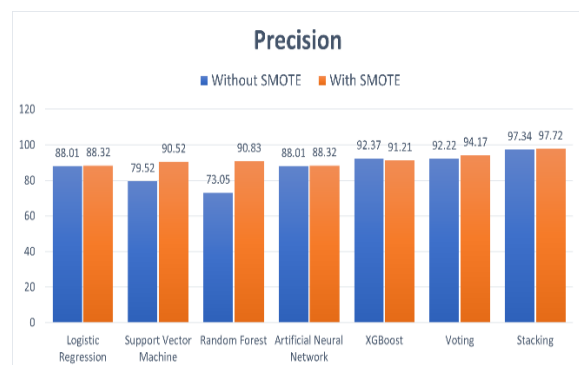


Figure 5. Evaluation of Model Precision Before and After Applying the SMOTE Technique

Precision is the most common and intuitive evaluation, representing the exact ratio of class predictions. The higher the precision value, the better the classification model built. Figure 5 shows that applying the SMOTE technique to imbalanced data, can improve the classification of data in the minority class and produces good precision in the majority class.

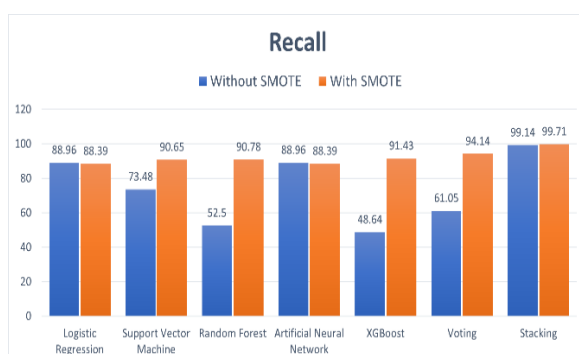


Figure 6. Evaluation of Model Recall Before and After Applying the SMOTE Technique

Recall or sensitivity is a model measurement technique in marking positive class samples which, if classified, will obtain results that are appropriate to the actual number of positive samples. The difference between precision and recall is that precision shows the system

to find the right one, while recall shows the system to find the perfect one. Figure 6 shows that the effect of the SMOTE technique can improve the evaluation of recall. Both evaluations will have high results when using ideal data, but under normal circumstances, recall will decrease as precision increases and vice versa. Therefore, precision and recall affect each other.

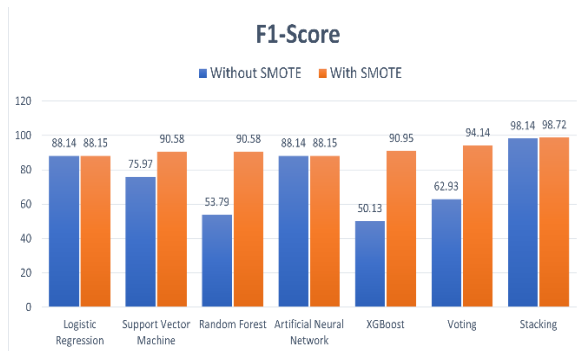


Figure 7. Evaluation of Model F1-Score Before and After Applying the SMOTE Technique

F1-Score value in Figure 7 proves that the SMOTE technique affects the accuracy of a model in predicting class. The lowest F1-Score value is 50.13%, this is because the model training uses an imbalanced dataset. The low value of the F1-Score in several individual models can also affect the performance of the hybrid method. Voting has a lower F1-Score than the other three individual methods, such as Logistic Regression, SVM, and ANN. This happens because two models have a very low F1-Score value so when combined it affects the performance of the Voting model. After using the SMOTE technique, the F1-Score value of each model increased, especially in the Voting method. When the data is combined with the SMOTE technique, the comparison between classes is balanced. Therefore, the more training data generated, the more accurate the model will be in guessing the prediction class.

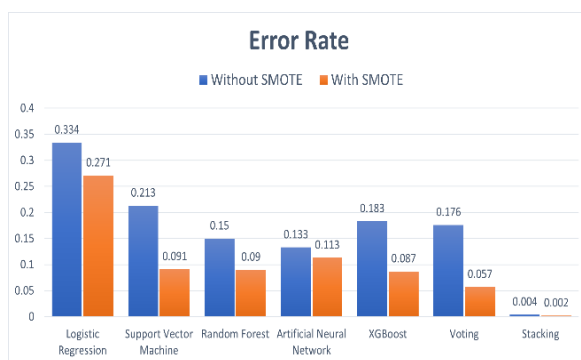


Figure 8. Comparison of Error Rate Before and After Applying the SMOTE Technique

Figure 8 shows the comparison of the error rate values generated by each classification model on imbalanced and balanced data. Error rate analysis illustrates that the lower the error rate value in the model, the better. The Hybrid method trains many models to correct the error value of each classification model and predicts the

output based on the highest probability of selecting a class as output, to produce the best accuracy.

Based on the results of the analysis of performance metrics using the hybrid Voting and Stacking methods, with 5 machine learning methods. Stacking can produce a higher model performance than Voting even though it uses imbalanced data types, by obtaining an accuracy value of 99.60%, precision 97.34%, recall 99.14%, and F1-Score 98.14%. While Voting obtained an accuracy value of 82.37%, precision 92.22%, recall, 61.05%, and f1-score 62.93%. When using the data with the influence of the SMOTE technique, the performance of the two hybrid methods gets better as the value of the evaluation model increases. Stacking results in an accuracy value of 99.71%, precision 97.72%, recall 97.71%, and F1-Score 98.72%. While Voting obtained an accuracy value of 94.30%, precision 94.17%, recall 94.14%, and f1-score 94.14%. The concept of hybrid Voting and Stacking method is to combine several individual classification models. However, both have differences in generating predictions, where Voting will choose the most effective classifier from each model while Stacking chooses the base classifier and overcomes imbalanced data. The two hybrid methods proposed in this study have better model performance than studies [12], [13], [15], [17], and [18]. This is because the Stacking model in this study obtained a model evaluation value of more than 97% in the test scenario using imbalanced and balanced data. While the results of the evaluation of the models in research [12], [13], and [15] were less than 97%. In addition to stacking, the performance of the Voting model with the influence of the SMOTE technique can produce an overall evaluation score of more than 94%. The voting model in this study is better than research [17] and [18] which only obtained an evaluation score of less than 94%.

#### 4. Conclusion

The performance of the model without SMOTE gives poor results with the accuracy of each individual classification model less than 90%. The combination method obtains an accuracy rate of 82.37% for Voting and 99.60% for Stacking. While by applying the SMOTE technique, the performance of single classification methods has increased to more than 90%, such as the SVM, Random Forest, and XGBoost methods. Then, the performance of the Voting method has succeeded in increasing and outperforming other individual classification models, especially at the level of accuracy that reaches more than 94%, while stacking reaches 99.71%. In analyzing the results of the comparison of the performance of the hybrid Voting and Stacking methods on rainfall classification, it can be concluded that by combining 5 machine learning methods on the data affected by the SMOTE technique, it can overcome the class imbalance problem in the data



and improve the performance of the classification model. These results confirm that the hybrid classification proposed using Stacking approach under the influence of the SMOTE technique, can accurately classify rainfall in the city of Bandung with an accuracy rate of 99.71% and almost reaches an excellent value of 100%.

### Acknowledgment

The author would like to thank Telkom University and the Bandung City Meteorology, Climatology and Geophysics Agency for the support of facilities and infrastructure in providing information and data, so that this research can be completed.

### Reference

- [1] BMKG, "Probabilistik Curah Hujan 20 mm (tiap 24 jam)," *Badan Meteorologi Klimatologi dan Geofisika*, 2022. <https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg> (accessed Jun. 13, 2022).
- [2] S. Ashilah, "Data Curah Hujan di Kota Bandung 1998-2020, Tahun 2010 Paling Basah," *BandungBergerak.id*, 2021. <https://bandungbergerak.id/article/detail/1351/data-curah-hujan-di-kota-bandung-1998-2020-tahun-2010-paling-basah> (accessed Jun. 02, 2022).
- [3] S. Ashilah, "Data Intensitas Banjir di Kota Bandung 2003-2020, Meninggi Seiring Fenomena Penurunan Muka Tanah," *BandungBergerak.id*, 2022. <https://bandungbergerak.id/article/detail/1606/data-intensitas-banjir-di-kota-bandung-2003-2020-meninggi-seiring-fenomena-penurunan-muka-tanah> (accessed May 15, 2022).
- [4] E. Kurniawan, F. Nhita, A. Aditsania, and D. Saepudin, "C5.0 algorithm and synthetic minority oversampling technique (SMOTE) for rainfall forecasting in Bandung regency," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, vol. 4, pp. 1–5, 2019, doi: 10.1109/ICoICT.2019.8835324.
- [5] N. A. Sulaiman, "Improving Support Vector Machine Rainfall Classification Accuracy based on Kernel Parameters Optimization for Statistical Downscaling Approach," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1.4, pp. 652–657, 2020, doi: 10.30534/ijatcse/2020/9191.42020.
- [6] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaiddm.v1i1.4903.
- [7] N. Oswal, "Predicting Rainfall using Machine Learning Techniques," 2019, doi: 10.36227/techrxiv.14398304.
- [8] I. Muslim and K. Karo, "Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan," *J. Softw. Eng. Inf. Commun. Technol.*, vol. 1, no. 1, pp. 10–16, 2020.
- [9] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," *J. Mater. Chem. C*, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
- [10] S. Shofura, S. Suryani M.Si, L. Salma, and S. Harini, "The Effect of Number of Factors and Data on Monthly Weather Classification Performance Using Artificial Neural Networks," *Int. J. Inf. Commun. Technol.*, vol. 7, no. 2, pp. 23–35, 2021, doi: 10.21108/ijoict.v7i2.602.
- [11] M. A. Rahman and R. C. Muniyandi, "Feature selection from colon cancer dataset for cancer classification using Artificial Neural Network," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, pp. 1387–1393, 2018, doi: 10.18517/ijaseit.8.4-2.6790.
- [12] D. D. Sidik and T. W. Sen, "Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan," *IT Soc.*, vol. 4, no. 1, pp. 21–27, 2019, doi: 10.33021/itfs.v4i1.1180.
- [13] S. K. Kalagotla, S. V. Gangashetty, and K. Giridhar, "A novel stacking technique for prediction of diabetes," *Comput. Biol. Med.*, vol. 135, no. June, p. 104554, 2021, doi: 10.1016/j.compbmed.2021.104554.
- [14] S. Gupta and M. K. Gupta, "Computational Prediction of Cervical Cancer Diagnosis Using Ensemble-Based Classification Algorithm," *Comput. J.*, vol. 00, no. 00, 2021, doi: 10.1093/comjnl/bxaa198.
- [15] S. Hou, Y. Liu, and Q. Yang, "Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning," *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 1, pp. 123–143, 2022, doi: 10.1016/j.jrmge.2021.05.004.
- [16] A. Nurmasani and Y. Pristyanto, "Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class," *Pseudocode*, vol. 8, no. 1, pp. 21–26, 2021, doi: 10.33369/pseudocode.8.1.21-26.
- [17] I. M. K. Karo, R. Ramdhani, A. W. Ramadhelza, and B. Z. Aufa, "A Hybrid Classification Based on Machine Learning Classifiers to Predict Smart Indonesia Program," *Proceeding - 2020 3rd Int. Conf. Vocat. Educ. Electr. Eng. Strength. Framew. Soc. 5.0 through Innov. Educ. Electr. Eng. Informatics Eng. ICVEE 2020*, 2020, doi: 10.1109/ICVEE50212.2020.9243195.
- [18] F. N. Khasanah and F. Nhita, "Weather Forecasting in Bandung Regency based on FP-Growth Algorithm," *Int. J. Inf. Commun. Technol.*, vol. 4, no. 2, p. 1, 2019, doi: 10.21108/ijoict.2018.42.203.
- [19] R. Ahuja and S. C. Sharma, "Stacking and voting ensemble methods fusion to evaluate instructor performance in higher education," *Int. J. Inf. Technol.*, vol. 13, no. 5, pp. 1721–1731, 2021, doi: 10.1007/s41870-021-00729-4.
- [20] A. Salam, Sri Suryani Prasetyowati, and Yuliant Sibaroni, "Prediction Vulnerability Level of Dengue Fever Using KNN and Random Forest," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, pp. 531–536, 2020, doi: 10.29207/resti.v4i3.1926.
- [21] H. Sain, H. Kuswanto, S. W. Purnami, and S. P. Rahayu, "Classification of rainfall data using support vector machine," *J. Phys. Conf. Ser.*, vol. 1763, no. 1, 2021, doi: 10.1088/1742-6596/1763/1/012048.
- [22] M. Y. H. Setyawan, R. M. Awangga, and S. R. Efendi, "Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot," *Proc. 2018 Int. Conf. Appl. Eng. ICAE 2018*, pp. 1–5, 2018, doi: 10.1109/INCAE.2018.8579372.
- [23] R. Gupta, N. Koli, N. Mahor, and N. Tejashri, "Performance analysis of machine learning classifier for predicting chronic kidney disease," *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–4, 2020, doi: 10.1109/INCET49848.2020.9154147.
- [24] S. H. Moon and Y. H. Kim, "An improved forecast of precipitation type using correlation-based feature selection and multinomial logistic regression," *Atmos. Res.*, vol. 240, no. February, p. 104928, 2020, doi: 10.1016/j.atmosres.2020.104928.
- [25] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005, doi: 10.1080/01431160412331269698.
- [26] N. Khalili, S. R. Khodashenas, K. Davary, M. M. Baygi, and F. Karimaldini, "Prediction of rainfall using artificial neural networks for synoptic station of Mashhad: a case study," *Arab. J. Geosci.*, vol. 9, no. 13, 2016, doi: 10.1007/s12517-016-2633-1.
- [27] A. Alamsyah and M. F. Permana, "Artificial Neural Network for Predicting Indonesian Economic Growth Using Macroeconomics Indicators," *Proceeding - 2018 Int. Symp. Adv. Intell. Informatics Revolutionize Intell. Informatics Spectr. Humanit. SAIN 2018*, pp. 15–19, 2019, doi: 10.1109/SAIN.2018.8673347.

- [28] Darussalam and G. Arief, "Jurnal Resti," *Resti*, vol. 1, no. 1, pp. 19–25, 2017.
- [29] D. N. Purba, Fhira Nhita, and Isman Kurniawan, "Implementation of Ensemble Method in Schizophrenia Identification Based on Microarray Data," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 64–69, 2022, doi: 10.29207/resti.v6i1.3788.
- [30] I. L. Cherif and A. Kortebi, "On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification," *IFIP Wirel. Days*, vol. 2019-April, pp. 1–6, 2019, doi: 10.1109/WD.2019.8734193.
- [31] S. Misra, H. Li, and J. He, *Machine Learning for Subsurface Characterization*. Elsevier Science, 2019.