Scientific Journal of Informatics
Vol. 12, No. 2, May 2025
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Comparison of Ensemble Forest-Based Methods Performance for Imbalanced Data Classification

Yunia Hasnataeni[1*], Asep Saefuddin[2], Agus M. Soleh[3]
[1, 2, 3]Department of Statistics and Data Science, IPB University, Indonesia

Abstract.
Purpose: Classification of imbalanced data presents a major challenge in meteorological studies, particularly in rainfall classification where extreme events occur infrequently. This research addresses the issue by evaluating ensemble learning models in handling imbalanced rainfall data in Bogor Regency, aiming to improve classification performance and model reliability for hydrometeorological risk mitigation.
Methods: Four ensemble methods: RF, RoF, DRF, and RoDRF were applied to rainfall classification using three resampling techniques: SMOTE, RUS, and SMOTE-RUS-NC. The data underwent preprocessing, stratified splitting, resampling, and 5-fold cross-validation. Performance was evaluated over 100 iterations using accuracy, precision, recall, and F1-score.
Result: The combination of DRF with SMOTE-RUS-NC yielded the most balanced results between accuracy (0.989) and computation time (107.28 seconds), while RoDRF with SMOTE achieved the highest overall performance with an accuracy of 0.991 but required a longer computation time (149.30 seconds). Feature importance analysis identified average humidity, maximum temperature, and minimum temperature as the most influential predictors of extreme rainfall.
Novelty: This research contributes a comprehensive comparison of ensemble forest-based methods for imbalanced rainfall data, revealing DRF-SMOTE as an optimal trade-off between performance and efficiency. The findings contribute to improved rainfall classification models and offer practical insight for disaster mitigation planning and resource management in tropical regions.

Keywords: Random forest-based methods, Imbalanced data, Resampling techniques, Ensemble learning, Rainfall classification
Received May 2025 / **Revised** May 2025 / **Accepted** May 2025

INTRODUCTION

Classification of imbalanced data is a widely studied problem, because when there is imbalance in the data, classification tends to produce biased models with almost zero sensitivity to minority classes. Even though none of the samples in the minority class are correctly classified, the accuracy can be almost perfect at 99%, this is because most of the majority class is correctly classified and the minority class is not taken into account. In other words, accuracy will not provide a clear and accurate picture of classification performance on imbalanced data [1]. The problem of data imbalance becomes even more crucial when applied in a real-world context, where data imbalance problems occur in various fields, such as in medical image data [2], stunting data [3], capital market data [4], bankruptcy risk data [5], IPB master program student study success data [6], and weather data [7]. Failure of the model to recognize patterns in minority classes can have serious consequences, as minority classes hold the most important information for decision-making [8]. This can lead to adverse strategic consequences ranging from wrong diagnosis to disaster mitigation failures, one of which can be caused by failing to classify rainfall properly.

Unbalanced rainfall data is often found in weather analysis, especially in tropical regions such as Indonesia. The imbalance between days with high, light, and no rainfall can lead to inaccurate classification models and this imbalance causes the classification model to be biased towards the majority class, resulting in low performance in recognizing minority classes, which can interfere with decision-making efforts in natural resource management, disaster mitigation, and more sustainable regional development [9]. One commonly used approach to overcome the problem of class imbalance in data is resampling techniques. There are two basic methods in resampling, namely undersampling and oversampling. Undersampling is done by

*Corresponding author.
Email addresses: yunia200698@gmail.com (Hasnataeni)
DOI: 10.15294/sji.v12i2.24269

randomly reducing the majority class, while oversampling is done by randomly generating artificial data for the minority class Resampling can be done in the form of undersampling the majority class, oversampling the minority class, or a combination of both [10]. In this research, the problem of unbalanced data will be handled by resampling using oversampling, undersampling, and a combination of both techniques. Using both resampling techniques allows exploration of a more holistic and comprehensive approach to overcoming data imbalance, providing opportunities to find optimal solutions that cannot be achieved with only one technique [11].

The oversampling technique that will be used in this research is called Synthetic Minority Oversampling Technique (SMOTE). This technique involves creating new samples from minority groups by analyzing and taking existing samples from the nearest minority group [4]. In contrast, Random Under-Sampling (RUS) will be used as an undersampling technique, which is the most commonly used method in undersampling techniques [12]. The selection of these two methods is based on the popularity of academic literature and empirical evidence. SMOTE is known to be effective in solving problems with imbalanced data with innovative methods to improve model accuracy [4], while RUS is an easy but effective method in reducing bias from the majority class of data so that model accuracy can be improved [12]. To maximize the results, this research will also integrate the NearMiss Cleaning (NC) method with the SMOTE and RUS methods. The purpose of integrating these three methods is to optimize the performance of imbalanced data handling, working by balancing the dataset by removing majority data points that are very close to the minority class to reduce noise and ensuring the model focuses on representative samples [13]. The use of hybrid methods such as SMOTE-RUS-NC is believed to improve classification accuracy and reduce overfitting and underfitting, especially on data with extreme imbalance classes [14].

Machine learning models perform better for analyzing unbalanced data than experimental tables and other statistical methods. This was proven in research conducted by Jeong et al. (2020), which showed that the application of machine learning algorithms can provide more accurate and stable classification results on datasets with unbalanced class distributions. However, challenges remain, especially in the case of a large number of classes and unbalanced data making it difficult to improve the accuracy of the model, so there is still great research potential in the development of machine learning models. In this regard, ensemble learning is a useful and growing approach [15]. Ensemble learning involves integrating different models to improve model accuracy [16]. Ensemble learning aims to achieve better prediction performance by reducing the noise or error between observed and predicted data. Ensemble learning methods are usually categorized into three, namely bootstrap aggregation (bagging), boosting, and stacking methods. All three categories attempt to fit their predictions to the observations by reducing model variance, bias, or both simultaneously. The main difference is that bagging and boosting usually work with homogeneous models, while stacking excels at combining heterogeneous models [17].

Approaches such as ensemble learning become particularly relevant in the context of complex data analysis, such as class imbalance, big data, and high noise [18]. One such significant class imbalance phenomenon is in precipitation data. This can be seen in the radar data-based precipitation classification research conducted by Putra et al. (2024) which shows that most observations fall in the light rainfall class or no rain, while moderate to heavy rainfall events are relatively rare and form a minority class, especially in tropical regions such as Indonesia [19]. Bogor Regency is one of the regions in Indonesia that has complex topographic conditions that have a great influence on rainfall fluctuations. Rainfall in Bogor Regency is a very crucial aspect that is difficult to predict and vulnerable to extreme weather events, which can affect various sectors such as agriculture, disaster mitigation, and water resource management [20]. With its varied topography and predominantly tropical climate, Bogor Regency often experiences significant fluctuations in rainfall, which can cause major impacts on food security and regional infrastructure. For example, heavy rainfall can increase the risk of flooding, potentially damaging agricultural land and disrupting people's economic activities [21]. Therefore, accurate rainfall classification is indispensable to support strategic decision-making in the region. With a proper understanding of rainfall patterns and intensity, authorities and stakeholders can design more effective interventions for disaster mitigation. Accurate and effective classification can not only identify extreme rainfall patterns but can also contribute to preparing appropriate preventive measures before disasters occur.

In this research, a Random Forest (RF) based method development will be used to classify rainfall data in Bogor Regency. This is based on various studies that found that Random Forest has proven to have good performance in various studies of unbalanced data. For example, in research conducted by Jasthy et al.

(2024), Random Forest showed the best performance with 80% accuracy in diagnosing lung disease using chest x-ray images (unbalanced data cases) [22]. Another research conducted by Das et al. (2023) showed that Random Forests are the best choice for credit card fraud detection (unbalanced data cases), with accuracy reaching 99.92% after cross-validation testing [23]. In another research conducted by Brown et al. (2012) also produced Random Forest as the best model from five unbalanced datasets [24]. Therefore, this research will compare several Random Forest developments, such as Rotation Forest (RoF), Double Random Forest (DRF), and Rotation Double Random Forest (RoDRF) which will be combined by applying the resampling method to improve the accuracy of rainfall classification in Bogor Regency. This comparison is expected to provide deeper insights into the most effective techniques to overcome classification challenges in unbalanced rainfall data.

METHODS

Data

The Meteorology, Climatology, and Geophysics Agency (BMKG) provided the data used in this research, which was accessed through the BMKG online data site. This data was obtained from the Citeko Meteorological Station in Bogor Regency, West Java Province, which includes daily climate measurement data for 2023. The data consists of ten variables, with rainfall as the response variable. This data was used to analyze various aspects relevant to the research, taking into account the unbalanced class composition of the response variable. The nine explanatory variables included can provide detailed information that will help in understanding and predicting the response variable more accurately. The variables used are described in detail in Table 1 below:

Table 1. Details of research variables

Variable	Description	Data Type
Y	Rainfall (RR)	Numeric (mm)
X1	Wind direction at maximum speed (ddd_x)	Categorical
X2	Most frequent wind direction (ddd_car)	Categorical
X3	Maximum wind speed (ff_x)	Numeric (m/s)
X4	Average wind speed (ff_avg)	Numeric (m/s)
X5	Average humidity (RH_avg)	Numeric (%)
X6	Duration of sunlight exposure (ss)	Numeric (hours)
X7	Maximum temperature (Tx)	Numeric (°C)
X8	Minimum temperature (Tn)	Numeric (°C)
X9	Average temperature (Tavg)	Numeric (°C)

Unbalanced data

Unbalanced data describes a situation where there is a significant difference in the number of examples between one or more classes in the data set. The class with the largest number of examples is referred to as the majority class, while the class with the smallest number of examples is referred to as the minority class [25]. Data imbalance can result in misclassification which affects the accuracy score and allows the minority class to be considered as outliers [26]. Unbalanced data can be overcome by resampling through various approaches, one of which is by reducing the sample in the majority class (undersampling) or adding samples to the minority class (oversampling) [4].

Random under-sampling (RUS)

Random under-sampling is a technique that randomly reduces the amount of data in the majority class until a balanced proportion of the minority class is reached. Although this method may change the distribution and representative characteristics of the majority class, and risks misclassification, RUS still shows competitive performance compared to other undersampling and data cleaning methods [27]. The following is an illustration of how RUS works [28].

Figure 1. Illustration of how RUS works

Synthetic minority oversampling technique (SMOTE)

Chawla et al. (2002) introduced SMOTE as an attempt to overcome data imbalance by resampling the data. The basic concept behind SMOTE aims to balance the number of examples between minority and majority classes by creating new data (synthetic data) using k-nearest neighbors [29]. The closest neighbor is calculated using the euclidean distance method, where the median value of the standard deviation of all numerical variables in the minority class is used as the difference in categorical variable values. The following is the euclidean distance formula [6]. The following is an illustration of how SMOTE works [28].

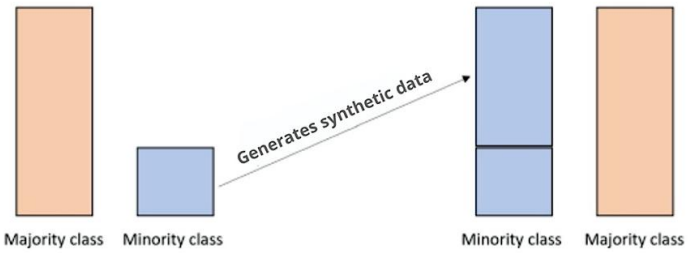

Figure 2. Illustration of how SMOTE works

SMOTE-RUS-nearmiss cleaning (NC)

The SMOTE-RUS-NC method combines three resampling techniques to handle data imbalance problems in classification, namely Synthetic Minority Oversampling Technique (SMOTE), Random Under-Sampling (RUS), and NearMiss Cleaning (NC). The NC method is used to solve the problem of class imbalance in the dataset by reducing the majority class that is very close to the minority class. The term "near-miss" refers to data points of the majority class that are very close to data points of the minority class, meaning they are almost similar, so they need to be removed to reduce the error in classifying the minority class [30]. When these three methods are integrated in a hybrid approach to improve dataset balance and classification model performance, especially on data with extreme class imbalance [14]. The following is an illustration of how SMOTE-RUS-NC works [28].

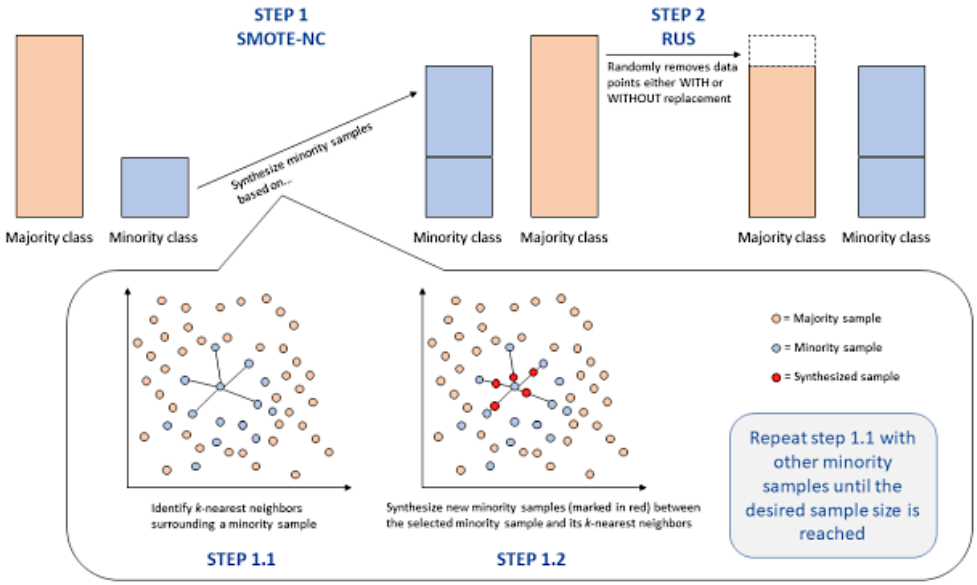

Figure 3. Illustration of how SMOTE-RUS-NC works

Random forest (RF)

Random Forest is one of the ensemble learning methods, developed by Leo Breiman in 2001. This method works by building a number of decision trees randomly from a subset of data with a bootstrap technique, each tree works independently to produce a model that is stable and able to handle data variations, then combines its prediction results through majority voting to improve accuracy and reduce overfitting [31]. The tree formation process starts from the root node to the leaves, with the best attribute at each node selected based on entropy and information gain calculations. This combination results in diversity between trees and produces more robust and accurate predictions [32].

$$Entropy\ (Y) = -\sum_i p(c|Y) log_2 p(c|Y) \quad (1)$$

Y : case set
$p(c|Y)$: proportion of Y value to class c

$$Information\ Gain\ (Y, a) = Entropy\ (Y) - \sum_{ve} values \frac{Y_v}{Y_a} Entropy\ (Y_v) \quad (2)$$

$values \frac{Y_v}{Y_a}$: all possible values in the set of cases a
Y_v : subclass of Y with class v yang corresponding to class a
Y_a : all values that correspond to a

$$Split\ Information\ (S, A) = \sum_i^c \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|} \quad (3)$$

$Split\ Information\ (S, A)$: estimated entropy value of input variable S case class c
$\frac{|S_i|}{|S|}$: probability of the i -th class in the attribute

$$Gain\ Ratio\ (S, A) = \frac{Information\ Gain\ (S, A)}{Split\ Information\ (S, A)} \quad (4)$$

Rotation forest (RoF)
The rotation forest algorithm introduced by Rodriguez et al. (2006) is a novel ensemble learning technique similar to random forest, where the training set for each base classifier is created through feature extraction. The main goal of rotation forest is to increase the diversity and accuracy of individuals simultaneously in the ensemble classifier [33]. Rotation forest is an ensemble classification method that uses principal component analysis (PCA) to rotate the variable axes before building a decision tree. The decision tree is chosen as the basis of classification due to its sensitivity to rotation of the variable axes as well as its ability to maintain accuracy. Despite using PCA, all principal components are still used in the construction of the decision tree to ensure the information remains complete. All estimation results from each classification tree are then combined using majority voting with the following equation [34].

$$\mu_j(x) = \frac{1}{B} \sum_{i=1}^{B} d_{i,j}(xR_i^a), \quad j = 1, \ldots, c \quad (5)$$

$\mu_j(x)$: The aggregate score for class j produced by the entire ensemble for input x
B : The number of trees (base classifiers) in the ensemble
R_i^a : The rotation matrix (resulting from PCA) used to transform input x for the i-th tree
c : The total number of classes

Double random forest (DRF)
Double Random Forest (DRF) is an ensemble learning method introduced by Liu and Wu in 2008. This method uses a combination of random instance bootstrapping and random feature subsets at each tree node. Unlike ordinary Random Forest which trains base models using bootstrap data, DRF trains each base model directly on the original data, resulting in more diverse features in the training process. This diversity allows the formation of more complex decision trees and improves generalization ability [35]. In addition, DRF also applies bootstrapping techniques to each non-terminal node. After the optimal features are selected from a subset of features randomly taken from the bootstrap sample, the original data is used for splitting. As a result, the original data is sent down the decision tree, resulting in more unique examples. This method applies randomization at two levels: the selection of training data subsets and feature subsets at each decision tree node, and the concatenation of multiple random forests, which together contribute to improving the accuracy and stability of the model [36].

Rotation double random forest (RoDRF)
Rotation Double Random Forest (RoDRF) was introduced by Ganaie et al. (2022) as a method that increases diversity between base models by randomly rotating data in different feature subspaces. The projections resulting from this transformation contribute to improved generalization ability and prediction accuracy.

Unlike conventional approaches, the variable rotation process is performed at each node using PCA, not just at the leaf level. In addition, this algorithm integrates the bagging principle at non-leaf nodes, thus enabling the formation of deeper trees to improve model performance [36].

Data analysis procedure

The data analysis procedure in this paper is illustrated in the flowchart in Figure 4 and consists of the following steps:

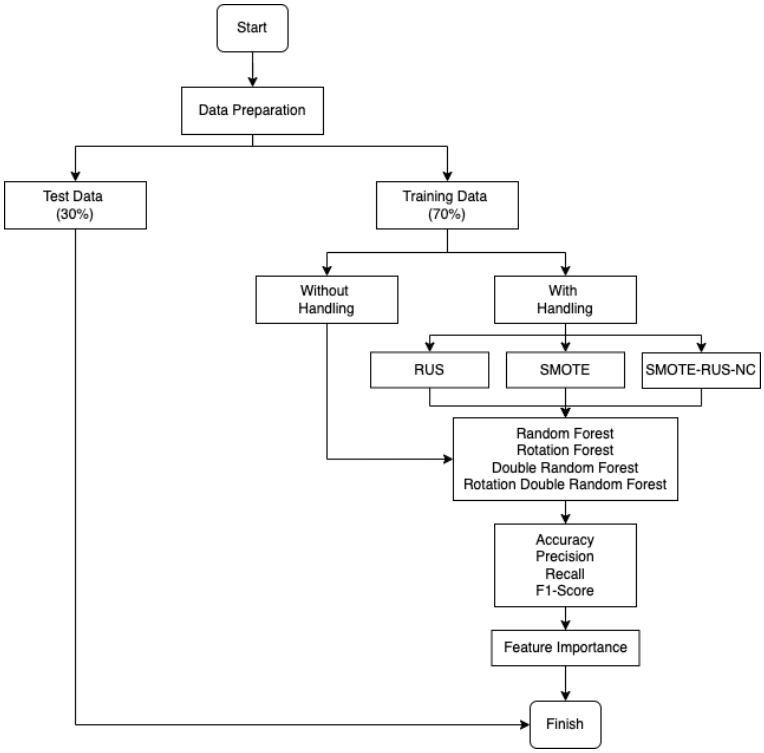

Figure 4. Flowchart of data analysis procedure

Flowchart of data analysis procedure:

1) Data Preparation: This stage includes exploratory data analysis through visualizations such as histograms, scatter plots, and heatmaps to understand distributions, correlations, and patterns in the data. Furthermore, data cleaning was carried out by removing missing, incomplete, or inaccurate data. In this process, a total of 25 entries with the code 8888 were identified, indicating missing values, along with 26 rows that were incomplete due to having at least one missing value. Next, data transformation was performed to adjust the format in accordance with the analytical methods to be used. For example, a categorical variable such as the most frequent wind direction (ddd_car) was converted into a numerical format using binary representation, namely 0 and 1. Finally, binarization of the target variable was conducted, where the rainfall variable (RR) was converted into two categories: extreme and non-extreme, based on thresholds established by BMKG. Rainfall values of 50 mm or more were categorized as extreme, while values below 50 mm were considered non-extreme.
2) Data Splitting: The data was split into 70% for training and 30% for testing using stratified sampling method to keep the proportion of target classes balanced.
3) Data imbalance handling: Data imbalance handling is done with three approaches: SMOTE, RUS, and a hybrid approach that combines RUS, SMOTE, and NearMiss Cleaning (NC).
4) K-Fold Cross-Validation: Model validation is performed with 5-fold cross-validation, which divides the training data into five subsets and trains the model alternately on each subset, while using the other subsets for validation.
5) Model Training: The model is trained using the ensemble methods of RF, RoF, DRF, and RpDRF.
6) Model Evaluation: Model performance evaluation is done by calculating metrics such as accuracy, precision, recall, and F1-score against test data to measure the classification ability of the model.

7) Iteration Process: The entire process from stage 1 to 6 is repeated 100 times to ensure the results obtained are consistent and reliable.
8) Significant Feature Identification: After the model evaluation is done, the analysis continues by identifying the features that have the most influence on extreme rainfall prediction through measuring feature importance.

RESULTS AND DISCUSSIONS

Exploration of meteorological data in Bogor Regency shows a relatively stable climate pattern with some indications of rare extreme events. The minimum temperature (Tn) is in the range of 14.8°C - 21.2°C (average 18.74°C), the maximum temperature (Tx) is between 20.2°C - 31.6°C (average 26.36°C), and the average temperature (Tavg) ranges from 18.8°C - 24.7°C (average 21.83°C), reflecting the stability of daily temperatures in this region. Average humidity (RH_avg) is quite high with a range of 58% - 100% and an average of 84%, which is characteristic of tropical regions. Rainfall (RR) varied significantly between 0 and 86.4 mm, but most days were dominated by light rainfall. After removing missing data (8888 mm values), the average rainfall was recorded at 6.99 mm. Extreme rainfall events (more than 50 mm) were very rare, as illustrated in Figure 5.

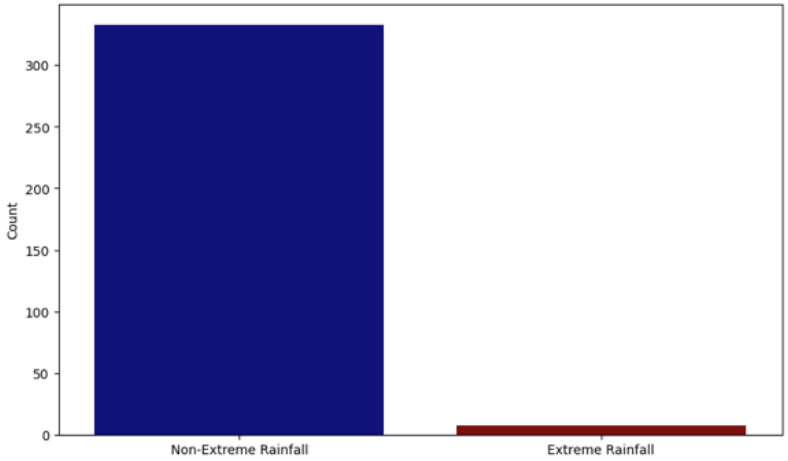

Figure 5. Rainfall distribution before handling class imbalance

The bar chart above shows an imbalanced rainfall distribution, with only 7 extreme rainfall instances compared to 333 non-extreme rainfall instances. This imbalance can lead to bias in classification models, making it necessary to apply data balancing techniques so that the model can recognize both classes proportionally. Figure 6 and 7 below presents the results of class balancing using SMOTE, RUS, and SMOTE-RUS-NC techniques.

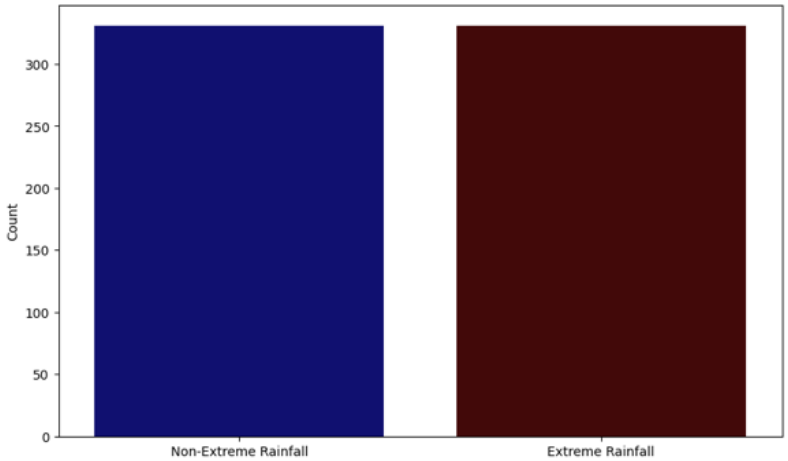

Figure 6. Rainfall distribution after handling with SMOTE and SMOTE-RUS-NC

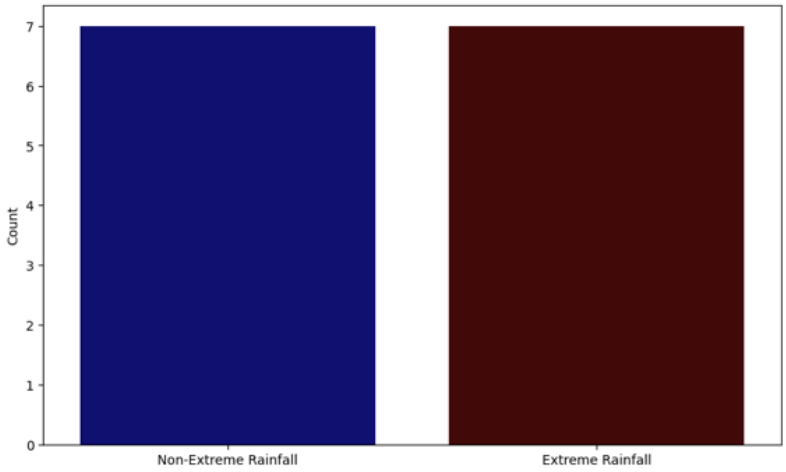

Figure 6. Rainfall distribution after handling with RUS

After handling class imbalance using SMOTE, RUS, and SMOTE-RUS-NC, the rainfall distribution became more balanced. SMOTE increased extreme rainfall instances to match the majority (331:331) by generating synthetic data. RUS reduced the majority class to match the minority (7:7) by randomly removing data. SMOTE-RUS-NC combined both approaches, resulting in a balanced dataset (331:331), offering a more robust solution for extreme imbalance. In Figure 7, the correlation between variables is examined to identify potential linear relationships among meteorological variables, including temperature, humidity, rainfall, wind speed, and solar radiation.

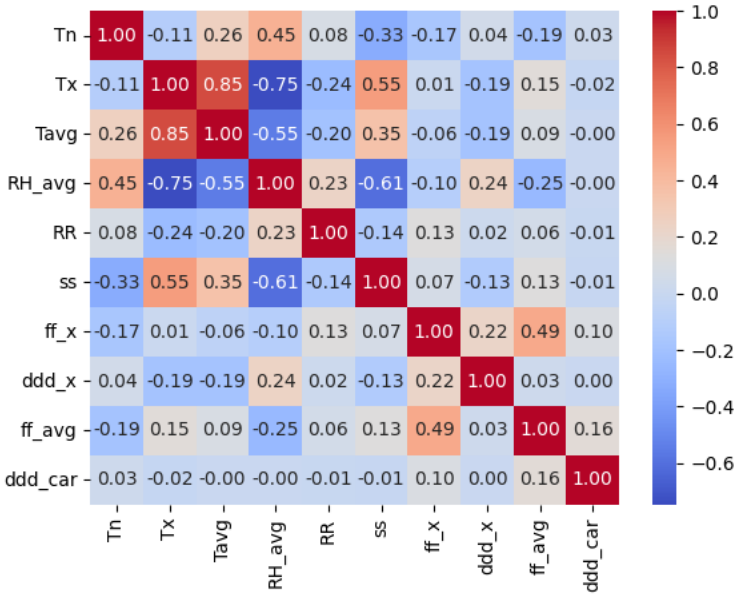

Figure 8. Correlation matrix between meteorological variables

Based on Figure 8, rainfall (RR) shows a weak correlation with other meteorological variables. The highest correlations are with maximum temperature (Tx) at -0.24 and average humidity (RH_avg) at 0.23, indicating that rainfall tends to decrease with higher temperatures and increase with higher humidity, although the relationships are not strong. The correlations between RR and other variables such as wind direction and sunlight duration are very low, suggesting that predictive models for extreme rainfall should not rely solely on linear relationships but also consider more complex non-linear approaches.

Overall, this exploration indicates a stable climate with the potential for occasional extreme weather events. Information on the relationship between temperature, humidity, rainfall, and wind speed can be utilized to understand local weather dynamics, improve the accuracy of weather predictions, and support mitigation efforts for hydrometeorological disasters such as floods and landslides in Bogor Regency. After

preprocessing, the dataset became cleaner and ready for analysis. Missing values (8888) and 26 incomplete rows were removed. Categorical variables like ddd_car was binarized, and rainfall (RR) was categorized based on BMKG's 50 mm threshold. Class imbalance was handled using SMOTE, RUS, and SMOTE-RUS-NC, resulting in balanced distributions. These steps prepared the data for fair and reliable model evaluation. Furthermore, to classify rainfall, the analysis was conducted using four ensemble learning methods: RF, RoF, DRF, dan RoDRF. This research compared the performance of each model in the face of unbalanced data, by assessing accuracy, precision, recall, and F1-score.

Without handling

In this section, four ensemble learning models are evaluated without any handling of data imbalance. Although all models show high accuracy, the performance of the models in classifying rainfall, especially in the minority class, is very low. The model performance results shown in the following table indicate the difficulty in distinguishing the majority and minority classes.

Table 2. Model performance without handling

Model	Accuracy	Precision	Recall	F1-Score	Time (Second)
RF_WH	0.983	0.177	0.167	0.169	104.44
RoF_WH	0.976	0.357	0.465	0.383	230.81
DRF_WH	0.981	0.078	0.072	0.074	79.88
RoDRF_WH	0.981	0.106	0.100	0.102	128.97

Based on the table above, all models show high accuracy (>0.97), with RF_WH recording the highest accuracy (0.983). However, the precision and recall values are very low, especially in DRF_WH, indicating a failure in recognizing minority classes. RoF_WH has the highest precision (0.357) and recall (0.465) among other models, indicating better performance in detecting extreme rainfall. The evaluation matrix visualization is presented in Figure 9 below.

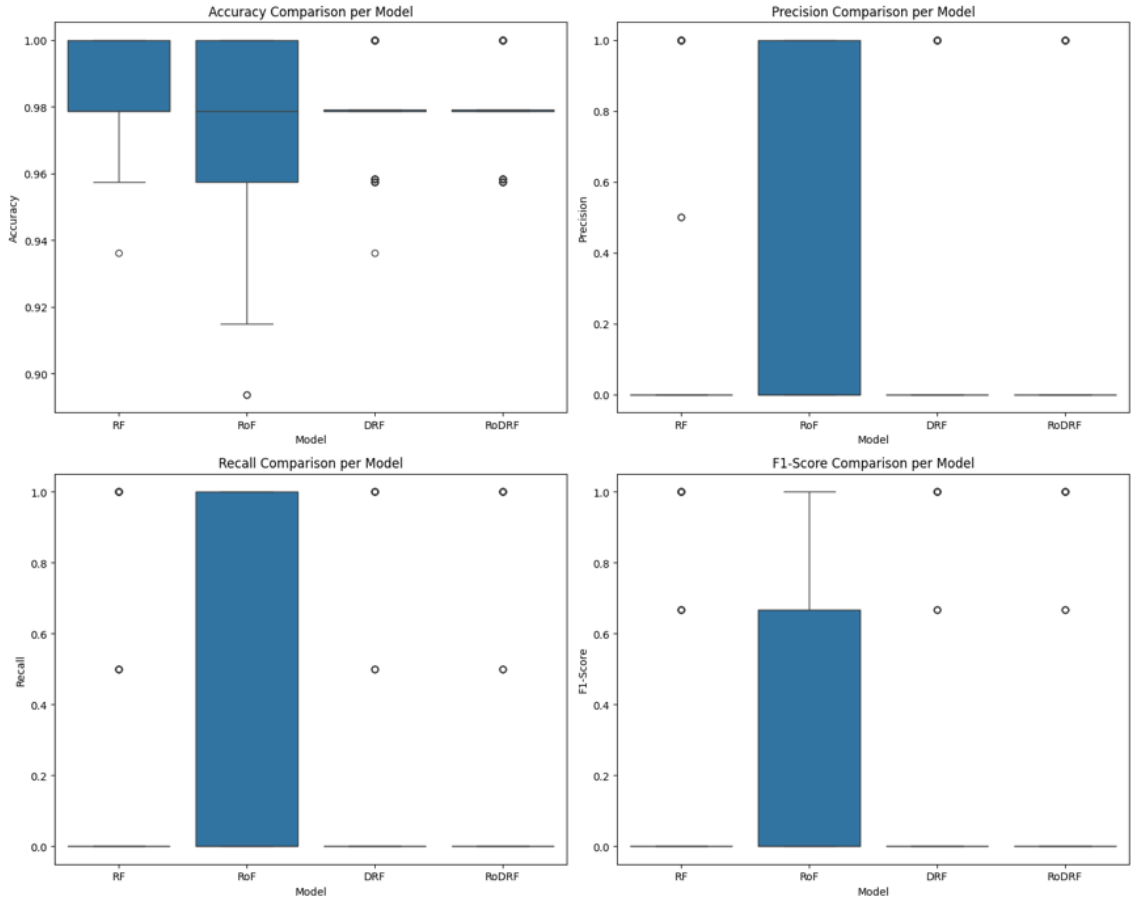

Figure 9. Boxplot of model performance without handlers

The boxplot shows that although RF_WH is stable in accuracy, this model like DRF_WH and RoDRF_WH tends to be biased towards the majority class. In contrast, RoF_WH performed more balanced in recognizing both classes. This finding confirms that handling data imbalance, such as SMOTE, is necessary to improve model performance in extreme rainfall classification.

Synthetic minority oversampling technique (SMOTE)

In this section, the SMOTE technique is applied to handle the imbalance of the data, after SMOTE is applied, the performance of the models has improved significantly. All models show high accuracy and are more balanced in recognizing majority and minority classes. The table below shows the performance of the models after the application of SMOTE.

Table 3. Model performance with SMOTE

Model	Accuracy	Precision	Recall	F1-Score	Time (Second)
RF_SMOTE	0.981	0.976	0.988	0.981	113.52
RoF_SMOTE	0.952	0.946	0.961	0.952	170.10
DRF_SMOTE	0.988	0.991	0.985	0.988	123.13
RoDRF_SMOTE	0.991	0.990	0.992	0.991	149.30

After the application of SMOTE, all models showed significant performance improvements. RoDRF_SMOTE recorded the best results with the highest accuracy, recall, and F1-score (all >0.99), demonstrating its ability to optimally handle data imbalance. DRF_SMOTE and RF_SMOTE also performed very well, while RoF_SMOTE although improved, remained slightly below the other models in some metrics. However, in selecting the best model, computation time is also an important consideration. RF_SMOTE is the model with the fastest computation time (113.52 second), followed by DRF_SMOTE, then RoDRF_SMOTE, and the longest is RoF_SMOTE. Therefore, although RoDRF_SMOTE gives the best results metrically, RF_SMOTE model can be a practical choice if time efficiency is a priority. The evaluation matrix visualization is presented in Figure 10 below.

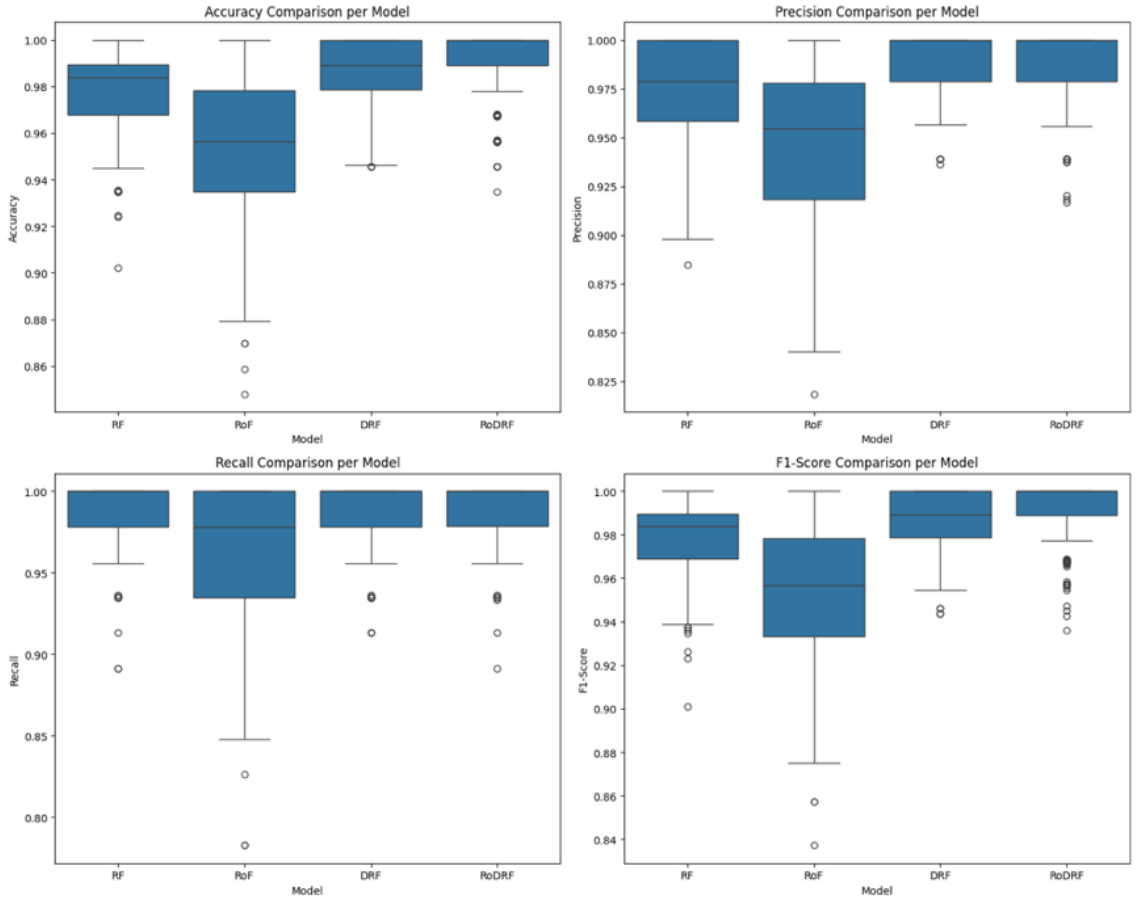

Figure 10. Boxplot of model performance with SMOTE

Figure 10 shows that the precision, recall, and F1-score distributions are more stable and higher in RoDRF_SMOTE and DRF_SMOTE, with fewer outliers. This shows the consistency of the model in recognizing majority and minority classes equally after the application of SMOTE.

Random under-sampling (RUS)

In applying the Random Under-Sampling (RUS) method, although there is an improvement in the balance between the majority and minority classes, the performance of the model decreases compared to the SMOTE treatment. However, it is still more stable than without handling. The following table shows the performance of the model after applying RUS.

Table 4. Model performance with RUS

Model	Accuracy	Precision	Recall	F1-Score	Time (Second)
RF_RUS	0.708	0.592	0.676	0.613	83.86
RoF_RUS	0.700	0.598	0.716	0.634	97.46
DRF_RUS	0.680	0.544	0.620	0.559	66.74
RoDRF_RUS	0.679	0.548	0.634	0.569	116.69

After the application of RUS, all models experienced a decrease in accuracy compared to SMOTE. RF_RUS recorded the highest accuracy (0.708), but low precision and recall, indicating an imbalance in recognizing minority classes. RoF_RUS had better precision and recall (0.598 and 0.716), although its accuracy was slightly lower (0.700). DRF_RUS and RoDRF_RUS showed the lowest performance, especially RoDRF_RUS with the lowest F1-score (0.569). The evaluation matrix visualization is presented in Figure 11 below.

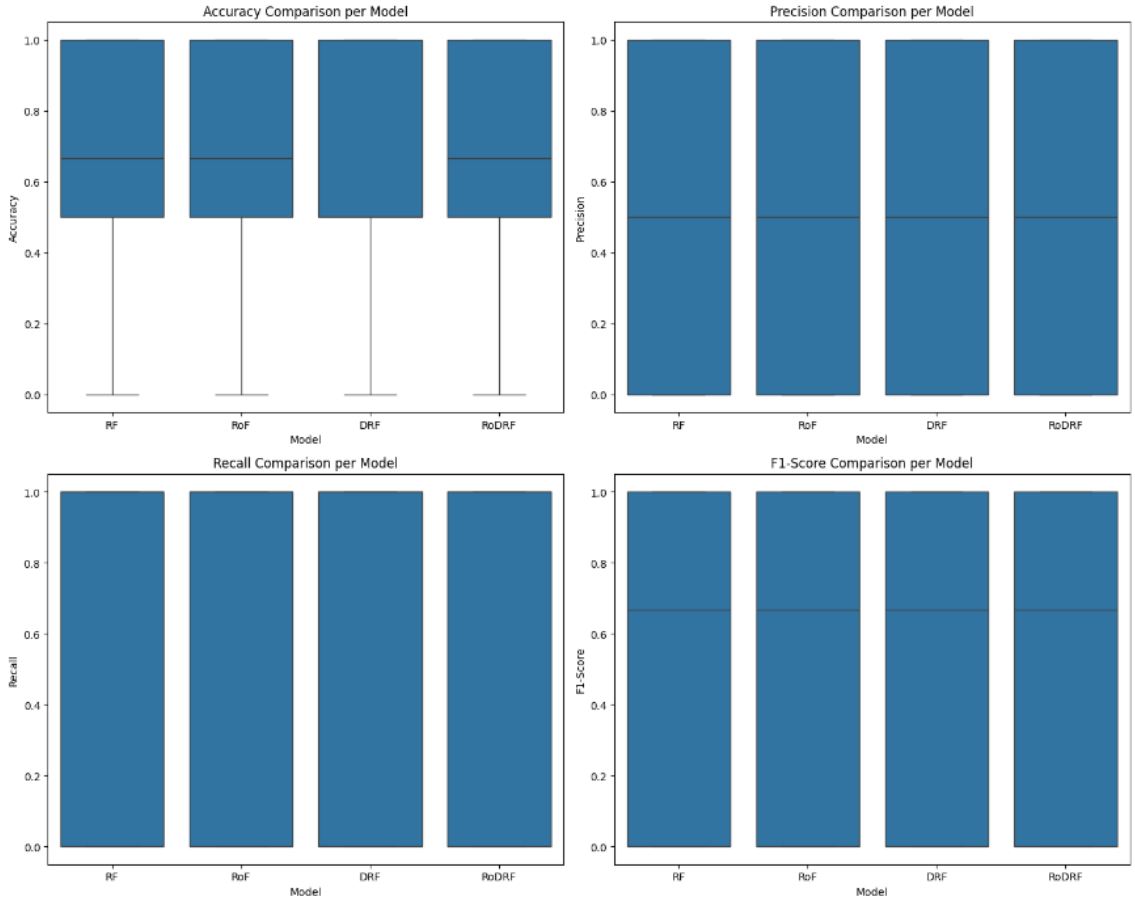

Figure 11. Boxplot of model performance with RUS

Figure 11 shows that no model consistently outperforms the others in terms of evaluation metric distribution. This indicates that RUS can help models better recognize minority classes compared to when

no imbalance handling is applied, although its performance is still less accurate and less optimal than the SMOTE method.

SMOTE-RUS-nearmiss cleaning (NC)

In this section, after applying the SMOTE-RUS-NC technique that combines SMOTE, RUS, and NC, the model performance shows significant improvement compared to the application of other techniques. This technique proves to be effective in handling data imbalance and improving the performance of the model in classifying rainfall. The table below shows the performance of the models after the application of SMOTE-RUS-NC.

Table 5. Model performance with SMOTE-RUS-NC

Model	Accuracy	Precision	Recall	F1-Score	Time (Second)
RF_SMOTE-RUS-NC	0.982	0.977	0.988	0.982	126.21
RoF_ SMOTE-RUS-NC	0.951	0.943	0.961	0.951	178.91
DRF_ SMOTE-RUS-NC	0.989	0.991	0.986	0.989	107.28
RoDRF_SMOTE-RUS-NC	0.989	0.988	0.991	0.989	168.39

The application of SMOTE-RUS-NC resulted in significant improvements in all models. DRF_SMOTE-RUS-NC and RoDRF_SMOTE-RUS-NC recorded high accuracy (0.989) along with excellent precision, recall, and F1-Score indicating the effectiveness of this technique in handling data imbalance better than using RUS. RoF_SMOTE-RUS-NC also performed well, but still below DRF and RoDRF. In considering the selection of the best model, computation time remains an important factor, DRF has the fastest execution time. Therefore, although DRF_SMOTE-RUS-NC and RoDRF_SMOTE-RUS-NC deliver nearly the same overall performance, DRF_SMOTE-RUS-NC can be considered an ideal alternative due to its balance between performance and computational efficiency. The evaluation matrix visualization is presented in Figure 12 below.

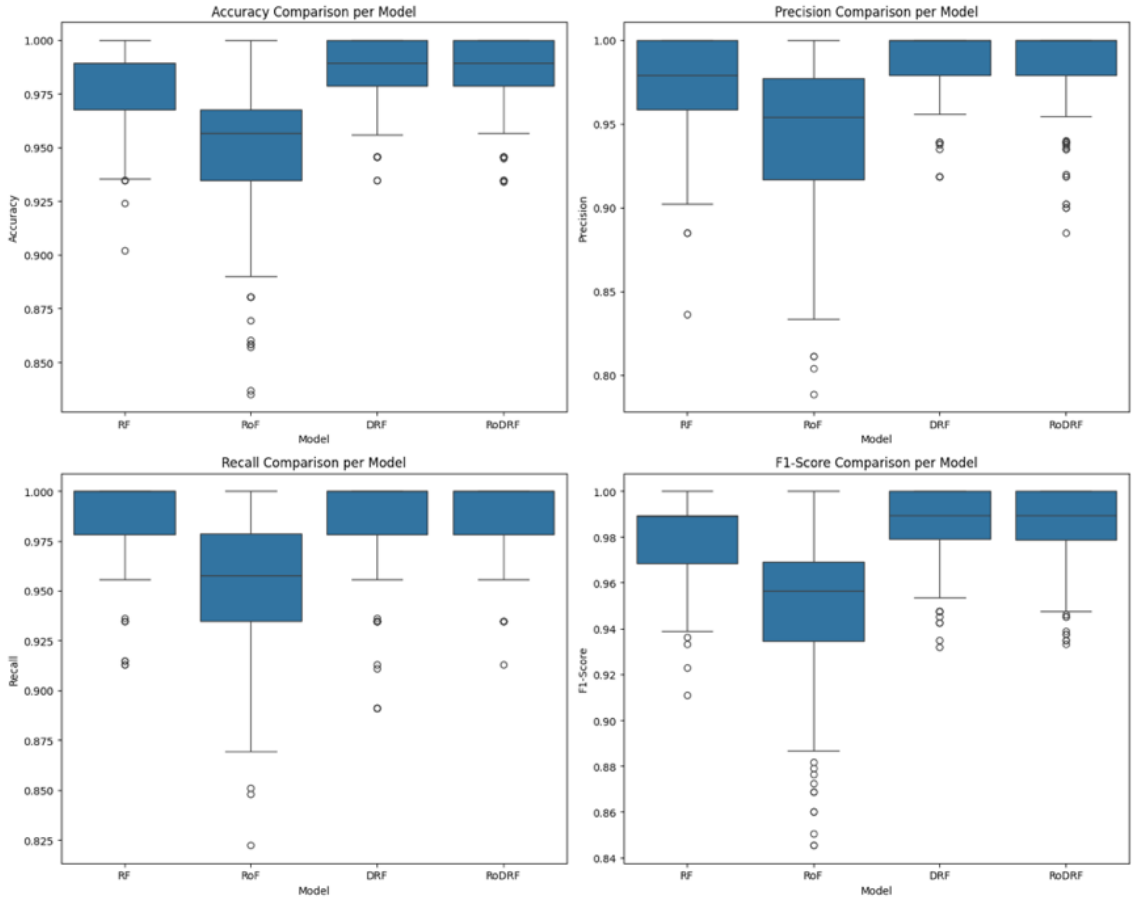

Figure 12. Boxplot of model performance with SMOTE-RUS-NC

Figure 12 shows a more stable and even distribution of metrics compared to RUS, especially in DRF and RoDRF, with smaller variations in precision and recall. This indicates that the model is able to recognize minority classes more consistently.

The best model

The boxplot in Figure 13 shows that models using the SMOTE and SMOTE-RUS-NC techniques (particularly DRF and RoDRF models) have high accuracy and stable distributions. The DRF_SMOTE and RoDRF_SMOTE models have accuracy levels that are not significantly different from those of the DRF_SMOTE-RUS-NC and RoDRF_SMOTE-RUS-NC models. However, DRF_SMOTE-RUS-NC is more efficient in terms of computation time. If time efficiency is a key consideration, DRF_SMOTE-RUS-NC is the most ideal choice, as it offers high accuracy with shorter computation time compared to other models. Overall, the SMOTE-RUS-NC technique has proven to be the most optimal in balancing model performance without sacrificing accuracy and efficiency.

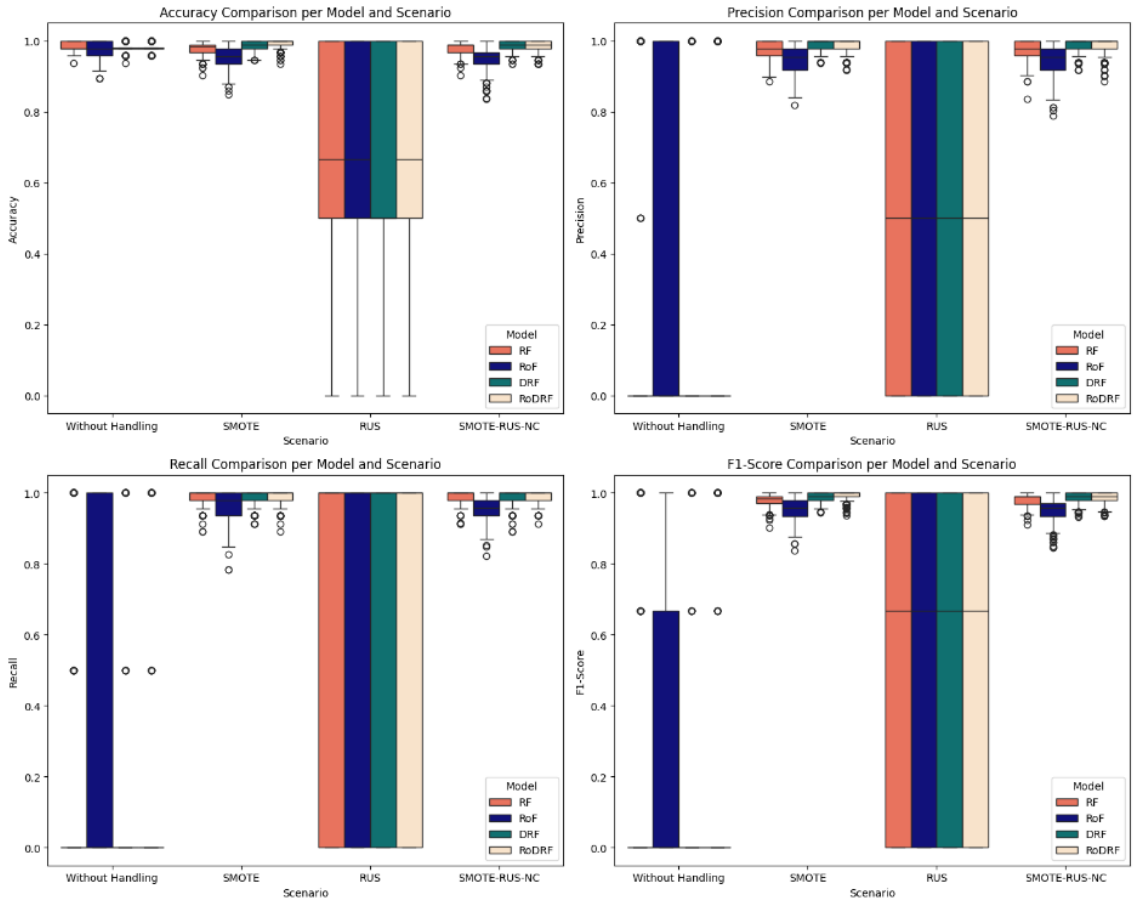

Figure 13. Boxplot of model performance

Feature importance of the best model

Figure 14 displays the feature importance of the SMOTE_DRF model. The most influential variable is RH_avg (average humidity), followed by Tx (maximum temperature), Tn (minimum temperature), and ss (sunshine duration). While variables such as Tavg, ff_x, and ddd_car have a lesser influence. The dominance of temperature and moisture variables in rainfall prediction indicates their importance in understanding local climate dynamics. This information is important for decision-making in disaster mitigation and natural resource management, especially in the face of potential climate change and extreme weather events.

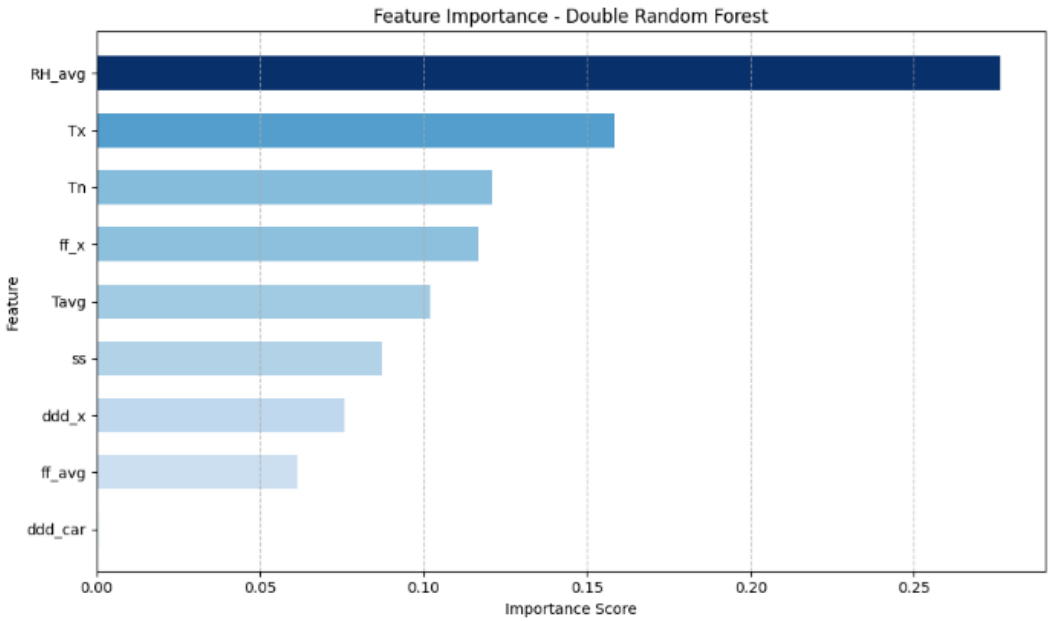

Figure 14. Feature Importance

CONCLUSION

The results of this research prove that handling unbalanced data is very important to get accurate classification results. The integration of ensemble learning and resampling techniques has proven to be highly effective in improving classification accuracy, particularly in detecting minority classes. Among the resampling methods, SMOTE and SMOTE-RUS-NC showed the best performance in balancing model sensitivity and specificity. The DRF model combined with SMOTE-RUS-NC provided a strong balance with an accuracy of 0.989 and a computation time of 107.28 seconds, while RoDRF with SMOTE yielded the highest average accuracy of 0.991 but required longer computation time (149.30 seconds). This trade-off can be considered when selecting a method based on specific priorities. In addition, average humidity, maximum temperature, and minimum temperature were identified as the most influential features for predicting rainfall intensity. Therefore, this research is expected to contribute to providing the best combination of methods for classifying extremely imbalanced rainfall data and serve as a reference for BMKG in designing and formulating policies.

REFERENCES

[1] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. Md Nasir, N. Shaadan, and S. W. Indratno, "Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 598–608, Jan. 2023, doi: 10.11591/ijeecs.v29.i1.pp598-608.

[2] J. X. Zhuang, J. Cai, J. Zhang, W. shi Zheng, and R. Wang, "Class attention to regions of lesion for imbalanced medical image recognition," *Neurocomputing*, vol. 555, Oct. 2023, doi: 10.1016/j.neucom.2023.126577.

[3] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data Optimizing Stunting Data Classification Through Ensemble Learning on Multiclass Labels with Imbalance Data."

[4] P. A. R. Mukhlashin, A. Fitrianto, A. M. Soleh, and W. Z. A. Wan Muhamad, "Ensemble learning with imbalanced data handling in the early detection of capital markets," *Journal of Accounting and Investment*, vol. 24, no. 2, pp. 600–617, May 2023, doi: 10.18196/jai.v24i2.17970.

[5] B. Siswoyo, Z. A. Abas, A. N. C. Pee, R. Komalasari, and N. Suyatna, "Ensemble machine learning algorithm optimization of bankruptcy prediction of bank," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 679–686, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp679-686.

[6] J. Wijaya, A. M. Soleh, and A. Rizki, "Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB," 2018.

[7] L. Mdegela, E. Municio, Y. De Bock, E. Luhanga, J. Leo, and E. Mannens, "Extreme Rainfall Event Classification Using Machine Learning for Kikuletwa River Floods," *Water (Switzerland)*, vol. 15, no. 6, Mar. 2023, doi: 10.3390/w15061021.

[8] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artif Intell Rev*, vol. 57, no. 6, Jun. 2024, doi: 10.1007/s10462-024-10759-6.

[9] H. Akbar and W. K. Sanjaya, "Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 3, no. 1, Dec. 2023, doi: 10.20885/snati.v3i1.30.

[10] M. S. Kraiem, F. Sánchez-Hernández, and M. N. Moreno-García, "Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association models," *Applied Sciences (Switzerland)*, vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188546.

[11] E. ismail, W. Gad, and M. Hashem, "SMOTE-RUS : Combined Oversampling and Undersampling Technique to Classify the Imbalanced Autism Spectrum disorder dataset," *International Journal of Intelligent Computing and Information Sciences*, vol. 23, no. 3, pp. 83–94, Sep. 2023, doi: 10.21608/ijicis.2023.216833.1278.

[12] S. Feng, J. Keung, Y. Xiao, P. Zhang, X. Yu, and X. Cao, "Improving the undersampling technique by optimizing the termination condition for software defect prediction," *Expert Syst Appl*, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121084.

[13] A. R. B. Alamsyah, S. R. Anisa, N. S. Belinda, and A. Setiawan, "SMOTE and Nearmiss Methods for Disease Classification with Unbalanced Data," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2021, no. 1, pp. 305–314, Jan. 2022, doi: 10.34123/icdsos.v2021i1.240.

[14] A. Newaz and F. S. Haq, "A Novel Hybrid Sampling Framework for Imbalanced Learning," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4200131.

[15] B. Jeong *et al.*, "Comparison between Statistical Models and Machine Learning Methods on Classification for Highly Imbalanced Multiclass Kidney Data," *Diagnostics*, vol. 10, no. 6, p. 415, Jun. 2020, doi: 10.3390/diagnostics10060415.

[16] M. Pirizadeh, N. Alemohammad, M. Manthouri, and M. Pirizadeh, "A new machine learning ensemble model for class imbalance problem of screening enhanced oil recovery methods," *J Pet Sci Eng*, vol. 198, p. 108214, Mar. 2021, doi: 10.1016/J.PETROL.2020.108214.

[17] Y. Zhang, J. Liu, and W. Shen, "A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications," *Applied Sciences*, vol. 12, no. 17, p. 8654, Aug. 2022, doi: 10.3390/app12178654.

[18] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst Appl*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.

[19] M. Putra, M. S. Rosid, and D. Handoko, "High-Resolution Rainfall Estimation Using Ensemble Learning Techniques and Multisensor Data Integration," *Sensors*, vol. 24, no. 15, p. 5030, Aug. 2024, doi: 10.3390/s24155030.

[20] Wiwik Tri Hardianti, "Penentuan Pola Distribusi Curah Hujan Harian Kabupaten Bogor dengan Model Rantai Markov," *Journal of Scientech Researchand Development*, vol. 6, no. 1, Jun. 2024.

[21] R. Hidayat and A. W. Farihah, "Identification of the changing air temperature and rainfall in Bogor," *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, vol. 10, no. 4, pp. 616–626, 2020, doi: 10.29244/jpsl.10.4.616-626.

[22] S. Jasthy, K. Ramasubramanian, R. Vangipuram, and S. Bollu, "Comparative Analysis of Machine-Learning Algorithms for Accurate Diagnosis of Lung Diseases Using Chest X-ray Images: A Study on Balanced and Unbalanced Data on Segmented and Unsegmented Images," *Cureus*, Jan. 2024, doi: 10.7759/cureus.53282.

[23] S. R. Das, R. Bin Sulaiman, and U. Butt, "Comparative Analysis of Machine Learning Algorithms for Credit Card Fraud Detection," 2023.

[24] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst Appl*, vol. 39, no. 3, pp. 3446–3453, Feb. 2012, doi: 10.1016/j.eswa.2011.09.033.

[25] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst Appl*, vol. 73, pp. 220–239, May 2017, doi: 10.1016/J.ESWA.2016.12.035.

[26] M. Koziarski, "Radial-Based Undersampling for imbalanced data classification," *Pattern Recognit*, vol. 102, Jun. 2020, doi: 10.1016/j.patcog.2020.107262.

[27] S. Ahmed, A. Mahbub, F. Rayhan, R. Jani, S. Shatabda, and D. M. Farid, "Hybrid Methods for Class Imbalance Learning Employing Bagging with Sampling Techniques," in *2nd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2017*, Institute of Electrical and Electronics Engineers Inc., Aug. 2018. doi: 10.1109/CSITSS.2017.8447799.

[28] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 1, p. 54, Jan. 2023, doi: 10.3390/info14010054.

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[30] A. Tanimoto, S. Yamada, T. Takenouchi, M. Sugiyama, and H. Kashima, "Improving imbalanced classification using near-miss instances," *Expert Syst Appl*, vol. 201, p. 117130, Sep. 2022, doi: 10.1016/j.eswa.2022.117130.

[31] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, "Efficient water quality prediction using supervised machine learning," *Water (Switzerland)*, vol. 11, no. 11, Nov. 2019, doi: 10.3390/w11112210.

[32] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 413–422, Feb. 2024, doi: 10.57152/malcom.v4i2.1210.

[33] T. Kavzoglu, I. Colkesen, and T. Yomralioglu, "Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image," *Remote Sensing Letters*, vol. 6, no. 11, pp. 834–843, Nov. 2015, doi: 10.1080/2150704X.2015.1084550.

[34] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A New classifier ensemble method," *IEEE Trans Pattern Anal Mach Intell*, vol. 28, no. 10, pp. 1619–1630, 2006, doi: 10.1109/TPAMI.2006.211.

[35] S. Han, H. Kim, and Y.-S. Lee, "Double random forest," *Mach Learn*, vol. 109, no. 8, pp. 1569–1586, Aug. 2020, doi: 10.1007/s10994-020-05889-1.

[36] M. A. Ganaie, M. Tanveer, P. N. Suganthan, and V. Snasel, "Oblique and rotation double random forest," *Neural Networks*, vol. 153, pp. 496–517, Sep. 2022, doi: 10.1016/j.neunet.2022.06.012.