JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 8, No. 1, Maret 2023, Pp. 115-126

PERFORMANCE ANALYSIS OF THE IMBALANCED DATA METHOD ON INCREASING THE CLASSIFICATION ACCURACY OF THE MACHINE LEARNING HYBRID METHOD

Azmi Aulia Rahman*1), Sri Suryani Prasetiyowati2), Yuliant Sibaroni3)
1. Telkom University, Bandung, Indonesia
2. Telkom University, Bandung, Indonesia
3. Telkom University, Bandung, Indonesia

Article Info
Keywords: Classification, Machine Learning, Hybrid Methods, Random Oversampling, Adasyn.

Article history:
Received 4 November 2022
Revised 18 November 2022
Accepted 25 November 2022
Available online 1 March 2023

DOI :
https://doi.org/10.29100/jipi.v8i1.3286

* Corresponding author.
Corresponding Author
E-mail address:
azmiar@student.telkomuniversity.ac.id

ABSTRACT
This study analyzes the performance of hybrid methods in improving accuracy on imbalanced data using Dengue Hemorrhagic Fever Case Data from 2017 to 2021 in Bandung City. The attributes used in this study consist of Total Population, Total Male, Elementary School Graduation, Junior High School Graduation, High School Graduation, College Graduation, Rainfall, Average Temperature, Humidity, Male Cases, Number of Cases, and Class. This research combines five Machine Learning methods, such as Decision Tree, Support Vector Machine, Artificial Neural Network, K-Nearest Neighbor, and Naïve Bayes. Hybrid Methods used in this research are Voting and Stacking methods. The oversampling methods used to handle imbalanced data in this study are Random Oversampling and Adasyn. The results show that Voting and Stacking without Random Oversampling and Adasyn get the same accuracy of 88,88%. While using Random Oversampling, voting gets an accuracy of 95,37% and stacking gets an accuracy of 96,29%. While using Adasyn, voting gets an accuracy of 94,44% and stacking gets an accuracy of 97,22%. Based on the results obtained, it can be concluded that the Random Oversampling and Adasyn Method can improve the performance of the Machine Learning hybrid method on imbalanced data. The contribution of this research is to provide information on the study and analysis of the implementation of the Random Oversampling and Adasyn methods in improving the performance of the Voting and Stacking methods in hybrid classification.

I. INTRODUCTION

CLASSIFICATION is a technique in machine learning using supervised learning algorithms that can predict the class of a new sample from a model inferred from training data, this classification classifies samples or examples of data into predetermined labels into a given set of data [1]. In this study, the authors used data on Dengue Hemorrhagic Fever (DHF) cases in Bandung City. Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus which is transmitted by mosquitoes of the Aedes Aegypti and Aedes Albopictus species [2]. Based on data from the Bandung City Health Office, dengue fever cases range from 2010 to 2018 the highest number of cases in 2013 was 5,736 cases while in 2014 it fell to 3,132, but again rose in 2015 to 3,640 cases and rose again in 2016 with a total of 3,880 cases, then decreased in 2017 by having a total of 1,786 cases and increased again in 2018 with a total of 2,826 cases [3]. In 2019 the number of cases reached 4,424 cases, compared to 2020 the number of Dengue Hemorrhagic Fever (DHF) cases in Bandung City decreased, namely in 2020 the number of cases was 2,790 cases. However, in 2021 there was an increase in the number of cases 2021 totaling 3,743 cases. In machine learning, unbalanced data is an important problem because unbalanced data can affect the model in the classification process. Unbalanced data occurs because the elements of a set of data are not evenly distributed or balanced across classes [4]. Unbalanced data on classification makes classification performance partial towards the majority class in an unbalanced data set, the majority class tends to lead towards solutions with better accuracy while the minority class gets poor accuracy results [5]. To handle the problem of unbalanced data in this study, oversampling can be used to balance the classes. Oversampling can solve problems in unbalanced data by increasing the number in the minority class [6].

Methods using Decision Tree and SVM were carried out in research [7]. This research discusses the prediction of Dengue Hemorrhagic Fever (DHF) with 18 attributes including classes. From the test accuracy results, the

Decision Tree method obtained an accuracy of 87,5% with a sensitivity of 90,9% and a specificity of 84,1%. Compared to the Decision Tree method, the SVM method gets a higher accuracy, which is 99%. The ANN method was used in research [8]. In research [8], the dataset used in this study consists of 110 instance data with sixteen attributes and has two classes as dengue positive and negative. The purpose of this study was to determine the relationship between Dengue diagnostic results, and environmental and physiological parameters. From the results of research conducted using the ANN method, the accuracy was 79,09% with a sensitivity of 55,55%, a specificity of 88,5%, and an error rate of 20,9%. In research [9] the K-NN method obtained an accuracy of 91%, precision of 90%, recall of 89%, and F1-Score of 89%. While Naïve Bayes got an accuracy of 93%, precision of 91%, recall of 90%, and F1-Score of 91%. The Naïve Bayes method was used in the research [10], in this study, the dataset used was collected from the 2016-2019 Dengue Hemorrhagic Fever (DHF) patient dataset in Semarang. The results of the research conducted, obtained by using the Naïve Bayes method without selection features on the training dataset obtained an accuracy of 67,4%, then on the testing dataset obtained an accuracy of 68,3%. While the testing process using feature selection on the training dataset, obtained an accuracy of 68%, then on the testing dataset, obtained an accuracy of 68,3%.

Classification with the Hybrid model was carried out in the research [2]. This research uses the Hybrid model in classifying Dengue Fever (DHF). Hybrid test results with the Voting method obtained an accuracy of 90%, a precision of 94%, recall of 82%, and F1-Score 86%. Combining Naïve Bayes and Decision Tree is done in research [11] combines Naïve Bayes and Decision Tree. From this research, it is analyzed that by using the Hybrid model, the accuracy can increase by 8% and get accuracy above 90%. Whereas if using only one method, the accuracy obtained is still below 90%. The Hybrid model in this study uses the Voting method for Hybrid classifiers. In research [12] the hybrid Voting method and applying oversampling to handle imbalanced class data, the hybrid Voting method gets an accuracy of 91%. Another hybrid method with the Stacking method was carried out in research [13] using a different hybrid method, namely the Stacking hybrid method, using Stacking with the Diabetes dataset to get an accuracy of 78,2%.

The use of the oversampling method in dealing with imbalanced data was done in the study [14] the oversampling technique was used, namely Random Oversampling. SVM, Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, and hybrid Voting methods were used in this study. In this study, there are two target classes in the dataset, namely 0 and 1. For the methods used, the SVM and Decision Tree methods apply hyperparameters with several different parameters. The SVM method with Linear kernel parameters gets an accuracy of 73%, SVM with RBF kernel gets an accuracy of 80%. The Naïve Bayes method gets an accuracy of 76%. Decision Tree method with default hyperparameter gets an accuracy of 94%, Decision Tree with hyperparameter criterion = entropy, max dept = 20 gets an accuracy of 87%, Decision Tree with hyperparameter = Gini, max depth = 18 gets an accuracy of 84%. The Logistic Regression method gets an accuracy of 73% and the Random Forest method with hyperparameter criterion = entropy, max depth = 25 gets an accuracy of 99,8% and the Hybrid Decision Tree method with parameters max depth = 18, Naïve Bayes, Logistic Regression, and Random Forest gets an accuracy of 93%.

The application of oversampling to deal with imbalanced data was conducted in the study [15][16] applied to oversampling to handle unbalanced data. In research [15] the hybrid method is used in this study, from the tests conducted, with imbalanced data the Majority Voting hybrid method gets an accuracy of 94%, Weighted Voting gets an accuracy of 94% and Stacking gets an accuracy of 95%, while using data with balanced, Majority Voting hybrid method gets an accuracy of 98,5%, Weighted Voting is 98,8% and Stacking is 99,6%. From these test results using ROS makes the data balanced and can increase accuracy. Research [16], shows that unbalanced data get an accuracy of 92% for the K-NN Method and 99,04% for the Naïve Bayes method. While with Random Oversampling K-NN accuracy gets 97% accuracy and Naïve Bayes gets 99,04% accuracy. Cases with imbalanced data are also used in [17] in this study using Smote in handling imbalanced data, from the test results obtained without Smote getting hybrid Voting accuracy of 82,37% and Stacking of 99,60% while with Smote hybrid Voting accuracy gets an accuracy of 94,30% and Stacking of 99,71%.

Other oversampling methods such as Adasyn can also handle imbalanced data. The use of Adasyn was carried out in research [18][19]. In research [18] using Statlog (Heart) Dataset, using the Naive Bayes method without oversampling got an accuracy of 85,2%, and using Adasyn got an accuracy of 84,3%. SVM method without oversampling gets an accuracy of 85,2% and Adasyn gets an accuracy of 84,3% and with Nearest Neighbor without oversampling, gets an accuracy of 83,3% and Adasyn gets an accuracy of 83,6%. Research [19] uses Adasyn to handle imbalanced data, the method used is SVM. From the test obtained an accuracy of 87,3%, this accuracy has increased from the accuracy without Adasyn, which is 83%. Research [20] uses oversampling methods to handle imbalanced data, one of the methods used is ANN with Adasyn. The feature selection techniques used are RFE and FCBF. In the original data, the accuracy is 0,6084%, while using Adasyn and RFE feature selection gets an accuracy of 0,6304%, and Adasyn and FCBF feature selection get an accuracy of 0,6467%.

Based on the research described above, taking into account the advantages and disadvantages carried out in previous studies. The author conducts research analyzing the performance of the hybrid Voting and Stacking method by combining five machine learning methods, such as Decision Tree, Support Vector Machine, Artificial Neural Network, K-Nearest Neighbor, and Naïve Bayes using Random Oversampling and Adasyn in handling imbalanced data using the Dengue Hemorrhagic Fever (DHF) dataset in Bandung City from 2017 to 2021.

II. RESEARCH METHODS

A. System Design

The system to be built is the classification of Hybrid Voting and Stacking methods with a combination of five machine learning methods Decision Tree, Support Vector Machine, Artificial Neural Network, K-Nearest Neighbor, and Naïve Bayes with three data usage scenarios, namely without Random Oversampling and Adasyn, with Random Oversampling, and with Adasyn.

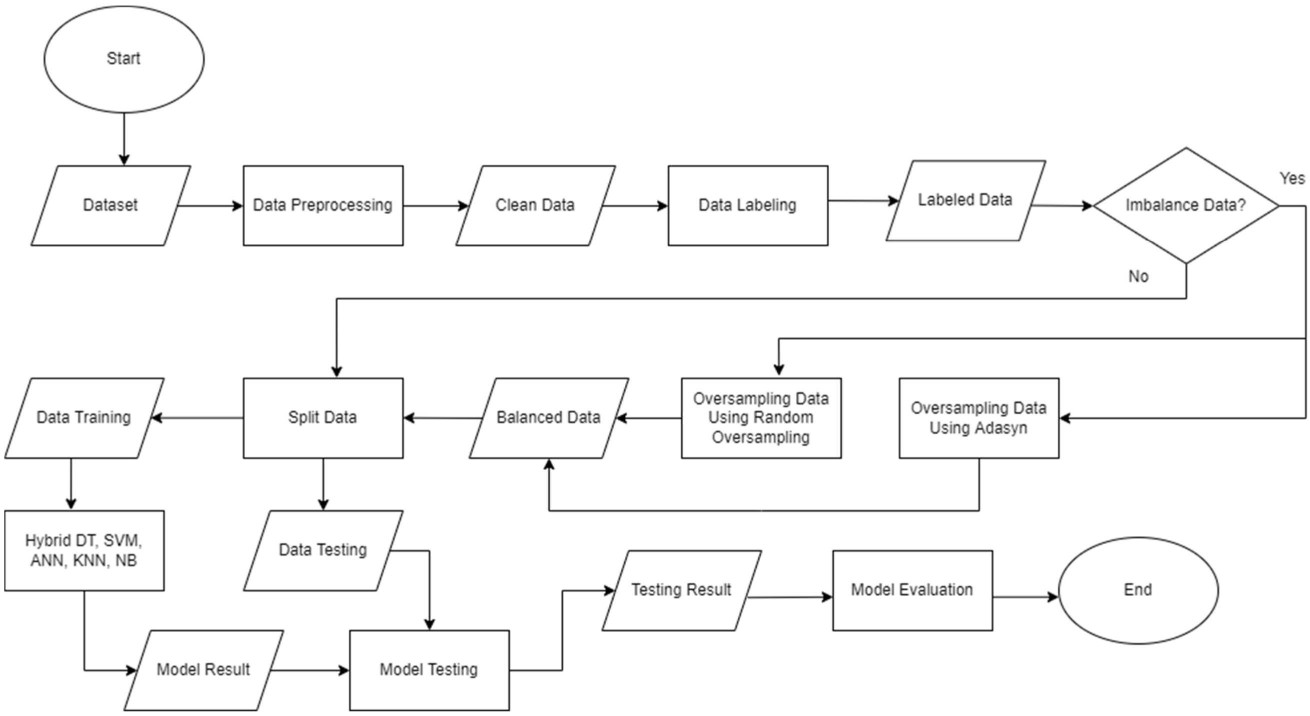

Figure 1. System Design Flowchart

B. Dataset

The dataset used is data on Dengue Hemorrhagic Fever cases based on sub-districts in Bandung City from 2017 to 2021. The attributes used in this research are Total Population, Total Male, Elementary School Graduation, Junior High School Graduation, High School Graduation, College Graduation, Rainfall, Average Temperature, Humidity, Male Cases, Number of Cases, and Class. The data came from the Bandung City Health Office, Central Statistics Agency, and Meteorology, Climatology, and Geophysics Agency.

TABLE I
SAMPLE DATA

No	Total Population	Total Male	Elementary School Graduation	Junior High School Graduation	High School Graduation	College Graduation	Rainfall (mm)	Average Temperature (°C)	Humidity (%)	Male Cases	Number of Cases
1	99085	49993	12323	15999	35744	51743	193,2	23,5	77,4	16	33
2	75209	37762	6154	7912	22661	30573	193,2	23,5	77,4	52	92
3	72424	36541	8127	9045	20255	29300	193,2	23,5	77,4	40	62
4	73236	36747	8677	11508	26368	37876	193,2	23,5	77,4	29	58
5	132497	67823	29951	22756	34170	56926	193,2	23,5	77,4	9	17
…	…	…	…	…	…	…	…	…	…	…	…
146	80808	40175	6853	9761	28615	38376	167,9	23,4	79	48	99
147	102766	51428	13472	14009	32551	46560	167,9	23,8	77	114	215
148	77601	38759	7845	8251	23148	31399	167,9	24,1	73	75	134
149	37921	18952	2872	4415	14487	18902	167,9	24,1	79	24	54
150	90006	45281	10354	12913	27544	40457	177,2	24,1	83	76	165

In Table 1, shows the sample data used in this study. The sample data used is 150 data from 30 sub-districts in Bandung City from 2017 to 2021.

C. Data Preprocessing

Preprocessing is the process of transforming raw data that has been collected into more efficient data by combining, selecting, and cleaning data so that the data can be processed easily.

D. Data Labeling

Data labeling is the process of identifying raw data and adding one or more important and informative labels to provide context so that machine learning models can learn from the process. Data labeling must be done correctly so that the data collected or classified according to the class that it should be. Based on research [21], which refers to the Epidemiological Window Bulletin with the main topic of Dengue Hemorrhagic Fever from the Epidemiological Data and Surveillance Center of the Ministry of Health of the Republic of Indonesia [22], there are three categories that represent the number of cases. The three categories are low, high, and medium. Of these categories, the number of cases less than 20 is categorized as low, the number of cases greater than or equal to 20 and less than or equal to 55 is categorized as a medium, and the number of cases greater than 55 is categorized as high. From research [2][12] related to Dengue Hemorrhagic Fever classification, the three categories are divided into class label 0, class label 1, and class label 2. Class label 0 is categorized as low class, class label 1 is categorized as a medium class, and class label 2 is categorized as high class.

TABLE II
CLASS LABELING

Class	Label Class	Range
Low	0	Number of Cases < 20
Medium	1	Number of Cases ≥ 20 and ≤ 55
High	2	Number of Cases > 55

E. Split Data

Split data is the process of dividing data into training data and testing data. This dataset is divided into two, namely training data and testing data with a ratio of 70:30. The division for the data, namely 70% is used for training data and 30% is used for testing data. To make the data have a range of 0 to 1, normalization is performed. For normalization, you can use MinMax Scaler [23]. The formula for MinMax Normalization is defined in equation (1).

$$X_{scaled} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

Where X_{min} is the minimum value in feature X and X_{max} is the maximum value in feature X [24].

F. Oversampling with Random Oversampling

The Oversampling method that is also used is to use Random Oversampling. Random Oversampling is a method that handles problems with unbalanced data in classification problems, this resampling process is carried out at the preprocessing stage [25]. This oversampling technique randomly selects minority samples to be copied to increase the proportion of minority classes [26]. The purpose of Random Oversampling is to balance the class distribution through random repetition of the minority class [14].

G. Oversampling with Adasyn

Adaptive Synthetic (Adasyn) is a method of oversampling that works by generating minority class instances [27]. The concept of Adasyn is to determine the weighted distribution of the minority sample with respect to the learning difficulty of the minority sample [28]. Adasyn generates synthetic minority class samples by focusing on samples that are more difficult to detect [29], samples that are difficult to be classified get higher weights, and generate more samples that have higher weights [30].

H. Classification Process

Classification is a subset of supervised learning that studies the mapping between inputs and outputs whose correct values are obtained from the supervisor, the training inputs will be assigned in classification into one of the specified classes [31]. In classification, the members of the data set will be classified by labels or categories and for new input examples, the classes or labels will be assigned to be predicted [32].

I. Decision Tree

The Decision Tree method is a machine learning method capable of working in recursive partitioning and can solve classification and regression problems [33]. This Decision Tree has a structure that includes root nodes, branches, and leaf nodes [34]. A Decision Tree is a tree that classifies instances by sorting them based on the values of features. For each node in the Decision Tree to represent a feature in the instance and each branch represents a value that can be estimated by that node to be classified, the instance is classified from the root node and sorted according to its feature value [35]. Decision trees are formed using entropy and information gain values [10]. The Decision Tree formula is as follows. Entropy Value Calculation. Entropy is a measure of uncertainty associated with random variables. Entropy will increase with increasing uncertainty or randomness [36]. The formula is calculated using equation 2.

$$\text{Entropy(D)} = \sum_{i=1}^{c} - p_i \log 2\, p_i \qquad (2)$$

Where p_i is the probability that an arbitrary tuple D belongs to class C and is estimated by $| C_{i,D} | / | D |$ [36].
Information Gain Calculation, information gain is the difference between the original information gain requirements (that is, based only on class proportions) and the new requirements (that is, obtained after partition A) [36]. As for the formula, it can be formulated as equation 3.

$$Gain(D, A) = \text{Entropy(D)} - \sum_{i=1}^{v} \frac{| D_v |}{| D |} Entropy(D_j) \qquad (3)$$

Where, D: a given data partition, A: attribute, v: partitioning tuples in D across multiple attributes of A that have different values of v, $| D_v |$ = number of samples for a given value of v, $| D |$ = number of all data samples. D is divided into v partitions or subsets, $\{D_1, D_2, ... D_j\}$ if D_j contains tuples in D that have a_j result from A [36].

J. Support Vector Machine

The Support Vector Machine (SVM) method is a machine learning method that can solve linear and non-linear problems and works well on practical problems [37]. SVM separates the attribute space with a hyper-plane to maximize the margin for different classes [10]. The concept of performing classification with the SVM algorithm is to find the best hyperplane and find the most optimal data separation space in different classes in the input space so that the kernel function and parameters used to affect the performance of the SVM model. To find the hyperplane can be measured by margin and find the maximum point, the hyperplane with the largest margin is found with the maximum marginal hyperplane (MMH) [10][38]. The margin is the distance between the hyperplane and the closest pattern of each class where the closest pattern is called the support vector [38]. SVM classification function can be calculated using equations 4 and 5.

$$\text{minimize } \frac{1}{2} ||\omega||^2 \qquad (4)$$

$$\text{subject to : } y_i\,(\omega .\, x_i - b) \geq 1, \forall x_i \qquad (5)$$

Where ω is the n-dimensional vector normal to the optimal hyperplane, b is the bias, and y_i is the classified variable in the set [39].

K. Artificial Neural Network

Artificial Neural Network (ANN) is a method that works by simulating or imitating the workings of the human brain to solve a problem, ANN consists of one or more input layers, hidden layers, and output layers. Each layer consists of many neurons and each neuron symbolizes a variable in the input layer [40]. The ANN method consists of processing units called neurons where these neurons have a function that can determine the activation of the neuron, the function is called activation which processes input signals that have been combined, then converts them into output signals [41]. To calculate the sum of the product weights $x_i w_{kj}$ (for i=0 to m) is usually denoted as netk as shown in equation 6 [42].

$$net_k = X_0 W_0 \sum_{i=1}^{m} X_i W_{kj} \qquad (6)$$

artificial neuron calculates the output y_k as a certain function of the net_k value defined in equation 7 [42].

$$y_k = f(net_k) \qquad (7)$$

Where x and y are input and output signals, w_{kj} synaptic weights, synapses, and f is the activation function [42].

L. K-Nearest Neighbor

The K-Nearest Neighbor (K-NN) method is one of the machine learning methods that can be used for classification, the classification is based on the closeness or similarity of the distance function [10][24]. In K-NN the k value of the nearest neighbor which is the number of nearest neighbors is used to classify data from the training data [10]. K-NN has a function to classify data based on training data taken from the k nearest neighbors, where k is the number of nearest neighbors [21]. Euclidean distance can be used to measure the nearest neighbor distance, as defined in equation 8.

$$\text{K-NN Euclidean distance } d_i = \sqrt{\sum_{i=1}^{p} (X_{2i} - y_{1i})^2} \qquad (8)$$

Where X_{1i} is sample data, X_{2i} is testing data, *i* is variable data, *d* is distance, and *p* is data dimension [38].

M. Naïve Bayes

Naïve Bayes is a probabilistic based classifier using Bayes Theorem, this method is based on conditional probabilities assuming independence between features, Naïve Bayes classifier assumes category class labels and categorizes data from the training set and values in the test data [43]. Bayes Theorem describes an event based on conditions that may be associated with the event [42]. The mathematical equation for Bayes' Theorem is defined in equation 9.

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} \qquad (9)$$

X and Y are represented as events, P(X) and P(Y) which represent the ratio of X and Y without regard to each other. P(X|Y) is the conditional probability of observing the occurrence of X if Y is true. P(Y|X) is the ratio of observing the specified occurrence of Y if X is true [42].

N. Hybrid

A Hybrid is a combination of two or more Classification methods, this method aims to build a highly accurate ensemble classifier by combining several individual classifiers to improve classification accuracy [44]. In this study, the Hybrid approach method was used, namely the Voting and Stacking methods. The Voting method is an ensemble technique in which a collection of classifiers are grouped together, each classification is collected based on the most votes from the basic set of classifiers [9]. The Voting method is performed by counting the votes received by each output class from individual classifiers and the final classification decision is considered to be the output class with the highest number of votes [44]. Stacking is a method that uses meta-learning algorithms to combine predictions from multiple base learning algorithms [13]. Stacking is used to improve accuracy by using a combination of base learning models because stacking can improve the performance of individual learning algorithms [13][45]. The stacking structure consists of two levels, level 0 and level 1 and the multiple base learners (level 0) are combined by the meta learner (level 1) [39].

O. Model Evaluation

Based on the model that has been built, an evaluation will be carried out to measure the performance of the built model. In this evaluation, accuracy, precision, recall, and F1-Score will be calculated. The performance will be calculated using the Confusion Matrix as defined in equations 10, 11, 12, and 13.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \text{ x } 100\% \qquad (10)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \text{ x } 100\% \qquad (11)$$

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 8, No. 1, Maret 2023, Pp. 000-000

$$Recall = \frac{True\ Positive}{True\ Positive + False\ negative} \text{ x } 100\% \quad (12)$$

$$F1 - Score = \frac{2(Precision \times Recall)}{Precision + Recall} \text{ x } 100\% \quad (13)$$

III. RESULT AND DISCUSSION

A. Data Collection

Based on the data labeling, there are 3 classes with class categories 0 (low), 1 (medium), and 2 (high) based on the number of cases. From the labeling on the original data as shown in table 2, it is found that the low category class with class 0 has a total of 3, the medium category with class 1 has a total of 27 and the high category with class 2 has a total of 120.

TABLE III
ORIGINAL DATA

Class	Label Class	Total
Low	0	3
Medium	1	27
High	2	120

From the data in Table 3, showing the comparison of the number of unbalanced classes, to deal with the unbalanced data between the classes, the Random Oversampling and Adasyn methods are used.

TABLE IV
DATA AFTER RANDOM OVERSAMPLING AND ADASYN

Class	Label Class	With Random Oversampling	With Adasyn
Low	0	120	120
Medium	1	120	120
High	2	120	120

Based on Table 4, with Random Oversampling and Adasyn, the number of classes 0, 1, and 2 have the same number of classes, which is 120. This number of classes adjusts and equates to the number of classes in the majority class.

B. Result

From the results of the tests carried out, the following is a comparison of accuracy, precision, recall, and F1-Score from the Decision Tree, SVM, ANN, K-NN, Naïve Bayes, Voting, and Stacking methods without Random Oversampling and Adasyn, with Random Oversampling, and with Adasyn.

TABLE V
ACCURACY WITHOUT RANDOM OVERSAMPLING AND ADASYN, WITH RANDOM OVERSAMPLING, AND WITH ADASYN

Method	Accuracy		
	Without Random Oversampling and Adasyn	With Random Oversampling	With Adasyn
Decision Tree	82.22%	80.55%	76.85%
Support Vector Machine	82.22%	87.96%	91.66%
Artificial Neural Network	80.00%	85.18%	88.88%
K-Nearest Neighbor	75.55%	89.81%	87.03%
Naïve Bayes	91.11%	83.33%	76.85%
Voting	88.88%	95.37%	94.44%
Stacking	88.88%	96.29%	97.22%

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 8, No. 1, Maret 2023, Pp. 000-000

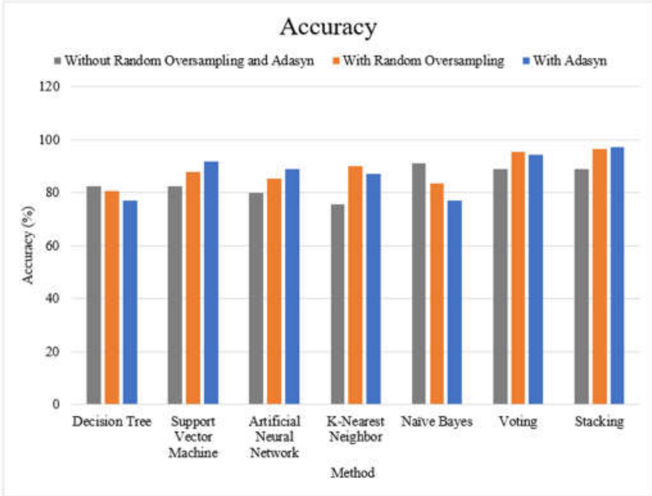

Figure 2. Accuracy Without Random Oversampling and Adasyn, With Random Oversampling, and With Adasyn

Based on Table 5 and Figure 2, showing tests conducted without Random Oversampling and Adasyn. Decision Tree and SVM methods have the same accuracy, which is 82,22%, while the ANN method gets an accuracy of 80% and the K-NN method gets an accuracy of 75,55%. Voting and Stacking methods also get the same accuracy, which is 88,88%. The Voting and Stacking methods without using oversampling have the same accuracy in predicting the classified model. The accuracy of the Voting and Stacking method is still higher than the other individual classification methods used in this study, such as Decision Tree, SVM, ANN, and K-NN, but the hybrid Voting and Stacking method are still below the accuracy of the Naïve Bayes method, where the Naïve Bayes method gets an accuracy of 91,11%. The use of Random Oversampling using the Voting and Stacking methods shows a significant increase in accuracy, using Random Oversampling, the hybrid Voting method gets an accuracy of 95,37% and Stacking gets an accuracy of 96,29%. The accuracy results with the hybrid Voting and Stacking method with Random Oversampling have increased compared to without using oversampling. Of the five individual classification methods combined, the K-NN method with Random Oversampling gets the highest accuracy for individual classification accuracy compared to other individual classification methods. Unlike Random Oversampling which randomly draws new samples to balance the minority class, the Adasyn method uses a weighted distribution to balance the data in each class. SVM, ANN, and K-NN methods with Adasyn have increased accuracy compared to using Random oversampling and Adasyn. When compared to without Random Oversampling and Adasyn, the hybrid Voting and Stacking method with Adasyn experienced a significant increase, while compared to Random Oversampling the accuracy of the hybrid Voting and Stacking method had a difference between Voting and Stacking accuracy between the two oversampling methods. In the Stacking hybrid method with Adasyn, the accuracy is 97,22%, the accuracy is increased compared to the accuracy of the stacking method with Random Oversampling, while the hybrid method with Voting using Adasyn gets an accuracy of 94,44%, the accuracy is still below the accuracy with the Voting hybrid method using Random Oversampling.

TABLE VI
PRECISION WITHOUT RANDOM OVERSAMPLING AND ADASYN, WITH RANDOM OVERSAMPLING, AND WITH ADASYN

Method	Precision		
	Without Random Oversampling and Adasyn	With Random Oversampling	With Adasyn
Decision Tree	81.24%	84.37%	83.67%
Support Vector Machine	76.86%	88.59%	92.12%
Artificial Neural Network	73.83%	86.84%	88.88%
K-Nearest Neighbor	72.22%	90.39%	87.19%
Naïve Bayes	86.92%	85.12%	80.17%
Voting	84.50%	95.38%	94.70%
Stacking	84.50%	96.38%	97.29%

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 8, No. 1, Maret 2023, Pp. 000-000

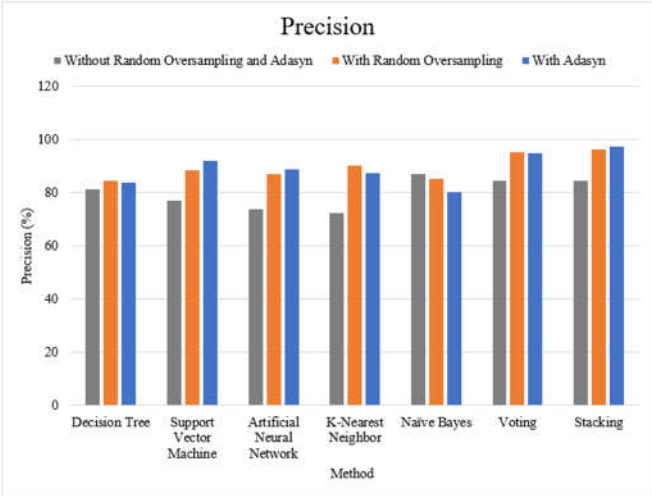

Figure 3. Precision Without Random Oversampling and Adasyn, With Random Oversampling, and With Adasyn

The precision score shown in Table 6 and Figure 3 shows the precision score with five individual classification methods and two hybrid methods have varying score values. Score precision is calculated based on the ratio of observations of positively correct predictions adjusted for the number of observations that can be predicted positively. The precision score of the hybrid voting and stacking method without using Random Oversampling and Adasyn gets the same precision score, which is 84,50%. Compared to using Random Oversampling, the hybrid Voting method gets a precision score of 95,38% and Stacking gets a precision of 96,38%, this precision increases compared to without Random Oversampling and Adasyn because from these tests imbalanced data and data with balanced can affect in building a classification model. In using Adasyn, the hybrid Voting and Stacking method has increased compared to the precision score without Random Oversampling and Adasyn, the hybrid Voting and Stacking method with Adasyn can increase the accuracy of the individual classification method, where Voting gets a precision of 94,70% and Stacking gets a precision score of 97,29%.

TABLE VII
RECALL WITHOUT RANDOM OVERSAMPLING AND ADASYN, WITH RANDOM OVERSAMPLING, AND WITH ADASYN

Method	Recall		
	Without Random Oversampling and Adasyn	With Random Oversampling	With Adasyn
Decision Tree	82.22%	80.55%	76.85%
Support Vector Machine	82.22%	87.96%	91.66%
Artificial Neural Network	80.00%	85.18%	88.88%
K-Nearest Neighbor	75.55%	89.81%	87.03%
Naïve Bayes	91.11%	83.33%	76.85%
Voting	88.88%	95.37%	94.44%
Stacking	88.88%	96.29%	97.22%

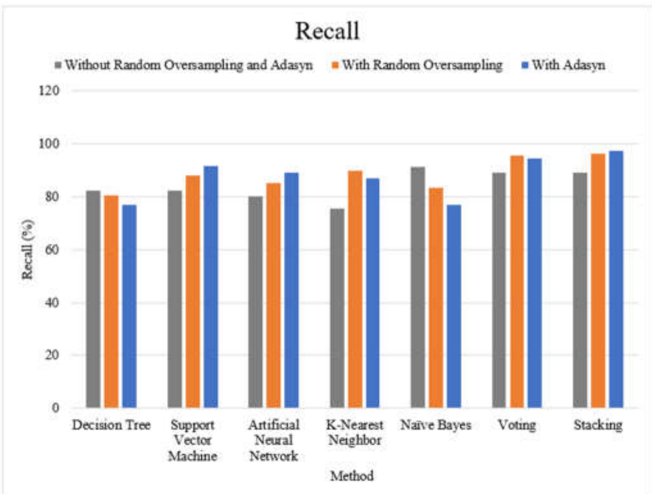

Figure 4. Recall Without Random Oversampling and Adasyn, With Random Oversampling, and With Adasyn

The recall score is shown in Table 7 and Figure 4, this recall is the ratio of positively predicted correct observations based on the total number of observations in the actual class. This recall is based on how exactly the success of the model was built in finding the data sought. For without Random Oversampling and Adasyn, the recall score with the Naïve Bayes method gets a recall score of 91,11%, and the recall score of the Naïve Bayes method is still higher than the hybrid Voting and Stacking method which gets the same recall score of 88,88%, with the hybrid Voting and Stacking method using without Random Oversampling and Adasyn has not been able to increase the recall score of the Naïve Bayes method but can increase the recall of other individual classification methods, this happens because the data used does not have a balanced number of class comparisons so as to make the performance of the Naïve Bayes method high. For Random Oversampling, the hybrid Voting and Stacking methods get a higher recall than other individual classification methods, where the Voting method gets a recall value of 95,37% and Stacking gets a recall of 96,29%. Adasyn can improve the hybrid Voting and Stacking method compared to without using Random Oversampling and Adasyn. The Voting method with Adasyn gets a recall value of 94,44% and Stacking with Adasyn gets a higher recall, which gets a recall value of 97,22%.

TABLE VIII
F1-SCORE WITHOUT RANDOM OVERSAMPLING AND ADASYN, WITH RANDOM OVERSAMPLING, AND WITH ADASYN

Method	F1-Score		
	Without Random Oversampling and Adasyn	With Random Oversampling	With Adasyn
Decision Tree	78.10%	80.26%	76.43%
Support Vector Machine	78.33%	87.84%	91.64%
Artificial Neural Network	75.19%	84.73%	88.59%
K-Nearest Neighbor	66.89%	89.61%	86.77%
Naïve Bayes	88.95%	82.19%	73.88%
Voting	86.51%	95.34%	94.43%
Stacking	86.51%	96.29%	97.22%

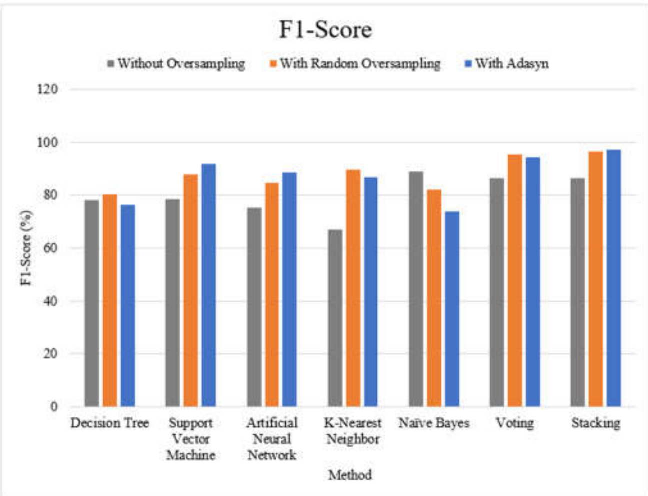

Figure 5. F1-Score Without Random Oversampling and Adasyn, With Random Oversampling, and With Adasyn

Based on Table 8 and Figure 5, without Random Oversampling and Adasyn, the Naïve Bayes method has a high accuracy compared to other individual classification methods, where the Naïve Bayes method gets an F1-Score of 88,95%. The hybrid Voting method gets an F1-Score of 86,51%, and Stacking gets an F1-Score of 86,51%. Combining the five combinations using the hybrid Voting and Stacking method, can improve four of the five individual classification methods used such as the Decision Tree method which gets an F1-Score of 78,10%, and an SVM of 78,33%, ANN of 75,19%, and K-NN of 66,89%. The hybrid voting and stacking method using Random Oversampling get the highest F1-Score score compared to other individual classification methods, this is because the hybrid Voting and Stacking method combines the five machine learning methods to get a higher F1-Score. The K-NN method with Random Oversampling gets the highest F1-Score for its individual classification method. By combining the five individual classification methods, the F1-Score of the hybrid Voting method is 95,34% and the F1-Score of Stacking is 96,29%. In Adasyn also the Hybrid Voting and Stacking methods have a high F1-Score improvement compared to without Random Oversampling and Adasyn. From the test results obtained, without Random Oversampling and Adasyn, using Random Oversampling, and using Adasyn, this SVM method with Adasyn is the method with the highest

F1-Score compared to all other individual classification methods. This shows that the SVM method using Adasyn is able to exceed the F1-Score of other individual classification methods. The hybrid Voting and Stacking method by combining five machine learning methods, such as Decision Tree, ANN, SVM, K-NN, and Naïve Bayes obtained an F1-Score of 94,43% for Voting and 97,22% for Stacking F1-Score. The test shows that using the hybrid Voting and Stacking method with Adasyn can increase the F1-Score even higher.

From the test results that have been carried out in this study, the hybrid voting method without oversampling has an accuracy below 90%, this accuracy is still below the accuracy of using the voting method in research [2], but when using balanced data gets a higher accuracy than research [2]. The Stacking method in this study without oversampling gets higher accuracy than in the research [13]. While using balanced data using the hybrid voting method by applying oversampling get higher accuracy than in research [12]. The use of Random Oversampling with the voting method in this study has higher accuracy than in research [14]. By using Adasyn using SVM and ANN methods in this study obtained higher accuracy than the research [18][19][20]. Tests conducted in this study using the hybrid Voting method with Random Oversampling and with Adasyn get higher accuracy than in research [17] which uses Smote to handle imbalanced data.

IV. CONCLUSION

From the tests carried out, imbalanced and balanced data can affect the performance of the classified model. Without using Random Oversampling and Adasyn, the hybrid Voting and Stacking method shows the same accuracy, which is 88,88%. Compared to using the Oversampling method with Random Oversampling and Adasyn makes the data balanced and increases the accuracy of the Hybrid method, with Random Oversampling, the Voting hybrid method gets an accuracy of 95,37% and Stacking gets an accuracy of 96,29%. While using Adasyn can increase the accuracy of the hybrid voting method by 94,44% and Stacking gets an accuracy of 97,22%. This show that using Random Oversampling and Adasyn method in handling imbalanced data in hybrid classification can improve the performance of the hybrid Voting and Stacking method.

ACKNOWLEDGMENT

The authors would like to thank Telkom University, Bandung City Health Office, Meteorology, Climatology and Geophysics Agency, and Central Statistics Agency for all their help and support in providing information and data needed in this research so that this research can be completed.

REFERENCES

[1] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0219-y.

[2] F. Nurul Inayah, S. Suryani Prasetiyowati, and Y. Sibaroni, "Classification of Dengue Hemorrhagic Fever (DHF) Spread in Bandung using Hybrid Naïve Bayes, K-Nearest Neighbor, and Artificial Neural Network Methods," *Int. J. Inf. Commun. Technol.*, vol. 7, no. 1, pp. 10–20, 2021, doi: 10.21108/ijoict.v7i1.562.

[3] M. Hasits, "Waspada, Awal Tahun 2019 Ada 48 Kasus DBD di Kota Bandung," 2019. https://bandung.merdeka.com/halo-bandung/read/164791/waspada-awal-tahun-2019-ada-48-kasus-dbd-di-kota-bandung (accessed Aug. 02, 2022).

[4] A. Fernández, S. del Río, N. V. Chawla, and F. Herrera, "An insight into imbalanced Big Data classification: outcomes and challenges," *Complex Intell. Syst.*, vol. 3, no. 2, pp. 105–120, 2017, doi: 10.1007/s40747-017-0037-9.

[5] H. Kaur, H. S. Pannu, and A. K. Malhi, "A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions," *ACM Comput. Surv.*, vol. 52, no. 4, 2019, doi: 10.1145/3343440.

[6] Y. T. Kim, D. K. Kim, H. Kim, and D. J. Kim, "A Comparison of Oversampling Methods for Constructing a Prognostic Model in the Patient with Heart Failure," *Int. Conf. ICT Converg.*, vol. 2020-Octob, pp. 379–383, 2020, doi: 10.1109/ICTC49870.2020.9289522.

[7] R. Sanjudevi and D. Savitha, "DENGUE FEVER PREDICTION USING CLASSIFICATION TECHNIQUES," *Int. Res. J. Eng. Technol.*, vol. 06, no. 02, pp. 558–563, 2019.

[8] S. Gambhir, S. K. Malik, and Y. Kumar, "PSO-ANN based diagnostic model for the early detection of dengue disease," *New Horizons Transl. Med.*, vol. 4, no. 1–4, pp. 1–8, 2017, doi: 10.1016/j.nhtm.2017.10.001.

[9] T. Sajana, M. Syamala, L. Phaneendra Maguluri, and C. Usha Kumari, "A hybrid approach for classification of infectious diseases," *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2020.11.727.

[10] A. Fahmi, D. Purwitasari, S. Sumpeno, and M. H. Purnomo, "Performance Evaluation of Classifiers for Predicting Infection Cases of Dengue Virus Based on Clinical Diagnosis Criteria," *IES 2020 - Int. Electron. Symp. Role Auton. Intell. Syst. Hum. Life Comf.*, pp. 456–462, 2020, doi: 10.1109/IES50839.2020.9231728.

[11] P. Taneja and N. Gautam, "Hybrid Classification Method for Dengue Prediction," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 1858–1861, 2019, doi: 10.35940/ijeat.F7892.088619.

[12] M. A. Rahman, S. S. Prasetiyowati, and Y. Sibaroni, "Performance Analysis of the Hybrid Voting Method on the Classification of the Number of Cases of Dengue Fever.," vol. 8, no. 1, pp. 10–19, 2022, doi: 10.21108/ijoict.v8i1.614.

[13] S. K. Kalagotla, S. V. Gangashetty, and K. Giridhar, "A novel stacking technique for prediction of diabetes," *Comput. Biol. Med.*, vol. 135, no. February, p. 104554, 2021, doi: 10.1016/j.compbiomed.2021.104554.

[14] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, pp. 243–248, 2020, doi: 10.1109/ICICS49469.2020.239556.

[15] S. Gupta and M. K. Gupta, "Computational Prediction of Cervical Cancer Diagnosis Using Ensemble-Based Classification Algorithm," *Comput. J.*, vol. 65, no. 6, pp. 1527–1539, 2022, doi: 10.1093/comjnl/bxaa198.

[16] M. Hayaty, S. Muthmainah, and S. M. Ghufran, "Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification," *Int. J. Artif. Intell. Res.*, vol. 4, no. 2, p. 86, 2021, doi: 10.29099/ijair.v4i2.152.
[17] A. Gumilar, S. S. Prasetiyowati, and Y. Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," vol. 5, no. 158, pp. 481–490, 2022.
[18] A. Gosain and S. Sardana, "Handling Class Imbalance Problem using Oversampling Techniques: A Review," *2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017*, vol. 2017-Janua, pp. 79–85, 2017, doi: 10.1109/ICACCI.2017.8125820.
[19] N. G. Ramadhan, "Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus," *Sci. J. Informatics*, vol. 8, no. 2, pp. 276–282, 2021, doi: 10.15294/sji.v8i2.32484.
[20] U. Ashfaq, D. B. P. M., and R. Mafas, "Managing Student Performance: A Predictive Analytics using Imbalanced Data," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 2277–2283, 2020, doi: 10.35940/ijrte.e7008.038620.
[21] A. Salam, Sri Suryani Prasetiyowati, 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.
[22] U. F. Achmadi *et al.*, "Buletin Jendela Epidemiologi Topik Utama Demam Berdarah Dengue," *Pusat Data dan Surveilans Epidemiologi Kementerian Kesehatan RI*, vol. 2, 2010.
[23] V. N. Jenipher and D. S. Radhika, "SVM kernel Methods with Data Normalization for Lung Cancer Survivability Prediction Application," *2021 Third Int. Conf. Intell. Commun. Technol. Virtual Mob. Networks*, pp. 1294–1299, 2021, doi: 10.1109/ICICV50876.2021.9388543.
[24] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020*, no. Icssit, pp. 729–735, 2020, doi: 10.1109/ICSSIT48917.2020.9214160.
[25] A. Fahmi, F. A. Muqtadiroh, D. Purwitasari, S. Sumpeno, and M. H. Purnomo, "A Multi-Class Classification of Dengue Infection Cases with Feature Selection in Imbalanced Clinical Diagnosis Data," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 3, p. 2022, 2022, doi: 10.22266/ijies2022.0630.15.
[26] W. Han, Z. Huang, S. Li, and Y. Jia, "Distribution-Sensitive Unbalanced Data Oversampling Method for Medical Diagnosis," *J. Med. Syst.*, vol. 43, no. 2, 2019, doi: 10.1007/s10916-018-1154-8.
[27] N. Rachburee and W. Punlumjeak, "Oversampling technique in student performance classification from engineering course," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, pp. 3567–3574, 2021, doi: 10.11591/ijece.v11i4.pp3567-3574.
[28] J. Liu, Y. Gao, and F. Hu, "A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM," *Comput. Secur.*, vol. 106, p. 102289, 2021, doi: 10.1016/j.cose.2021.102289.
[29] P. Gnip, L. Vokorokos, and P. Drotár, "Selective oversampling approach for strongly imbalanced data," *PeerJ Comput. Sci.*, vol. 7, pp. 1–22, 2021, doi: 10.7717/PEERJ-CS.604.
[30] M. S. Shelke, P. R. Deshmukh, and P. V. K. Shandilya, "A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique," *Int. J. Recent Trends Eng. Res.*, vol. 3, no. 4, pp. 444–449, 2017, doi: 10.23883/ijrter.2017.3168.0uwxm.
[31] A. E. Mohamed, "Comparative Study of Four Supervised Machine Learning Techniques for Classification," *Int. J. Appl. Sci. Technol.*, vol. 7, no. 2, 2017, [Online]. Available: www.ijastnet.com
[32] P. C. Sen, M. Hajra, and M. Ghosh, *Supervised Classification Algorithms in Machine Learning: A Survey and Review*, vol. 937. Springer Singapore, 2020. doi: 10.1007/978-981-13-7403-6_11.
[33] A. Z. Abdullah, B. Winarno, and D. R. S. Saputro, "The decision tree classification with C4.5 and C5.0 algorithm based on R to detect case fatality rate of dengue hemorrhagic fever in Indonesia," *J. Phys. Conf. Ser.*, vol. 1776, no. 1, pp. 0–10, 2021, doi: 10.1088/1742-6596/1776/1/012040.
[34] S. Pitchumani Angayarkanni, "Predictive Analytics of Chronic Kidney Disease using Machine Learning Algorithm," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 940–947, 2019, doi: 10.35940/ijrte.B1727.078219.
[35] O. F.Y, A. J.E.T, A. O, H. J. O, O. O, and A. J, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol.*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijctt-v48p126.
[36] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, "Analysis of Various Decision Tree Algorithms for Classification in Data Mining," *Int. J. Comput. Appl.*, vol. 163, no. 8, pp. 15–19, 2017, doi: 10.5120/ijca2017913660.
[37] K. T. Swe, P. Thu, and Z. Tun, "DENGUE FEVER CLASSIFICATION TOOL USING MACHINE LEARNING," vol. 2, no. May, pp. 5–10, 2020.
[38] D. A. Anggoro and N. D. Kurnia, "Comparison of Accuracy Level of Support Vector Machine (SVM) and K Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, pp. 1689–1694, 2020, doi: 10.30534/ijeter/2020/32852020.
[39] X. Hu, H. Zhang, H. Mei, D. Xiao, Y. Li, and M. Li, "Landslide susceptibility mapping using the stacking ensemble machine learning method in lushui, southwest China," *Appl. Sci.*, vol. 10, no. 11, 2020, doi: 10.3390/app10114016.
[40] M. Hayri Kesikoglu, U. Haluk Atasever, F. Dadaser-Celik, and C. Ozkan, "Performance of ANN, SVM and MLH techniques for land use/cover change detection at Sultan Marshes wetland, Turkey," *Water Sci. Technol.*, vol. 80, no. 3, pp. 466–477, 2019, doi: 10.2166/wst.2019.290.
[41] P. Silitonga, B. E. Dewi, A. Bustamam, and H. S. Al-Ash, "Evaluation of Dengue Model Performances Developed Using Artificial Neural Network and Random Forest Classifiers," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 135–143, 2021, doi: 10.1016/j.procs.2020.12.018.
[42] N. Iqbal and M. Islam, "Machine Learning for Dengue Outbreak Prediction: A Performance Evaluation of Different Prominent Classifiers," *Informatica*, vol. 43, no. 3, 2019, doi: 10.31449/inf.v43i3.1548.
[43] A. Prabhat and V. Khullar, "Sentiment classification on Big Data using Naïve Bayes and Logistic Regression," *2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017*, 2017, doi: 10.1109/ICCCI.2017.8117734.
[44] C. Qi and X. Tang, "A hybrid ensemble method for improved prediction of slope stability," *Int. J. Numer. Anal. Methods Geomech.*, vol. 42, no. 15, pp. 1823–1839, 2018, doi: 10.1002/nag.2834.
[45] N. Nahar, F. Ara, M. A. I. Neloy, V. Barua, M. S. Hossain, and K. Andersson, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction," *ICIET 2019 - 2nd Int. Conf. Innov. Eng. Technol.*, pp. 23–24, 2019, doi: 10.1109/ICIET48527.2019.9290507.