Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

Performance Analysis of Bandung City Traffic Flow Classification with Machine Learning and Kriging Interpolation

Nuraena Ramdani, Sri Suryani Prasetyowati*, Yuliant Sibaroni

School of Computing, Telkom University, Bandung, Indonesia
Email: [1]nuraenaramdani@student.telkomuniversity.ac.id, [2,*]srisuryani@telkomuniversity.ac.id, [3]yuliant@telkomuniversity.ac.id
Email Penulis Korespondensi: srisuryani@telkomuniversity.ac.idocc
Submitted:**25/07/2022**; Accepted:**18/08/2022**; Published**: 30/09/2022**

Abstract−This research focuses on making classification maps using the Classification And Regression Trees (CART), Random Forest and Ordinary Kriging methods. The dataset used is data from the Area Traffic Control System (ATCS) of the Bandung City Transportation Agency and the Google Maps application in April 2022. After the dataset is obtained, then the data pre-processing process will be carried out then the CART and Random Forest classification learning models will be made, after the CART and Random Forest classification learning is complete. From the CART and Random Forest classification models, traffic congestion classification map will then be made using the ArcMap application with the Ordinary Kriging interpolation method. The results of the comparison of classification maps made with Ordinary Kriging interpolation with the Gaussian Model semivariogram in both methods, namely CART and Random Forest. With the CART method has an accuracy of up to 88% while the classification map made with the Random Forest method has an accuracy of up to 90%. This proves that in this study the Random Forest method is far superior in building classification maps compared to the CART method.

Keywords: Classification Map; Traffic congestion; Random Forest; Cart; Ordinary Kriging

1. INTRODUCTION

Bandung City is the capital of West Java province and is one of the metropolitan cities in Indonesia with diverse activities such as in terms of economy, business, industry, education, and government and so on. Based on data from the Bandung City Statistics Agency (2020) in 2018 - 2020, there was an increase in population every year. In 2018 the total population in Bandung City is 2,503,708, experienced an increase in new residents of 4,180 so that the total population in Bandung City in 2019 was 2,507,888, and in 2020 also experienced an increase in population of 2,215 new residents so that in 2020 the total population in Bandung City was 2,510,103 [1]. This increase makes Bandung quite densely populated, and every resident must need a means of transportation to support their daily activities. According to the Head of the Bandung City Transportation Agency, EM Ricky Gustiadi (2021), revealed that the comparison of the use of private vehicles with public transportation in Bandung City is 80% of the population chooses to use private vehicles such as motorbikes and cars, while 20% of other residents choose to use public transportation such as angkot, online motorcycle taxis, and buses [4]. Therefore, due to the increase in population every year and the increase in private vehicle ownership, traffic congestion occurs [16]. Traffic congestion is a condition where a traffic flow is stopped due to obstruction of vehicle mobility which often occurs in big cities in developing countries such as Indonesia [2][3]. Usually, traffic jams in Bandung City often occur in the morning when leaving for school or work, in the afternoon during break time, and in the afternoon when leaving work.

The impacts caused by congestion are generally negative. Seen from various aspects such as material, time and health. As in the material or economic aspect, congestion results in the process of production and distribution rates being hampered so that the economy is disrupted, and not forgetting the waste of fuel which is detrimental to vehicle users on the road. Furthermore, the time aspect, congestion has an impact on wasting time which results in delays for someone to carry out activities such as school and work. And from the health aspect, congestion causes many negative things such as affecting the physical and psychological condition of road users so that emotions often occur between road users, and do not forget that air pollution and noise pollution generated from congestion are very bad for the body condition of road users [19].

Therefore, to solve problems related to traffic congestion in Bandung City and avoid the adverse effects of traffic congestion, the proposed solution in the field of informatics is a classification map of traffic congestion in Bandung City using the implementation of machine learning and Ordinary Kriging methods. Machine learning is an artificial intelligence approach that is widely used to be able to help problems that are difficult to solve by humans. machine learning algorithms can certainly solve these problems [20]. The selected machine learning implementation is using the Classification And Regression Trees (CART) method, Random Forest which is quite well-known and widely used to handle fairly accurate classification problems, so it is expected to produce high accuracy if the machine learning model runs with the correct process stages [5][9][14]. While Ordinary Kriging is an interpolation method that is often used by many people to predict values at surrounding location points where classification results are not available, interpolation data is obtained from the classification results that have been made [33].

There is previous research on Supervised learning, Supervised learning is the most common technique in classification techniques, because the goal is to make machines learn the classification system created. The learning process in machine learning models is divided into two steps, namely training and testing. In the training process, samples in the training data are taken as input in which features are learned with the learning algorithm and build the learning model. In the testing process, the learning model uses machine execution to make predictions for the test

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

data. The tagged data is the output of the learning model which gives the final prediction [6]. Data preprocessing is done to improve the quality of input data or to convert unstructured data into structured data.

Random Forest is one of the supervised learning classification methods that is quite well-known and widely used. Random Forest uses several decision trees to train data samples and integrate the weights of each tree to get the output that has the most votes as the final result [7][14]. Meanwhile, Classification and Regression Trees (CART) is one of the decision tree algorithms that combines classification trees and regression trees. The advantages of CART are mentioned that CART computation is faster [8].

There is research [23] that has been done before to find out the level of congestion in the city of Beijing, China by making a congestion level classification map. The result of this research is that the classification map created is an area zone classification map. The advantage of this research is that the classification map created has many types of maps. The disadvantage of this research is that it does not include the accuracy results of the classification. The difference between research [23] and this research is that the classification made only uses k-means clustering [23], while this research uses the CART and Random Forest methods.

Then in research [11] to overcome congestion using the Random Forest classification algorithm technique to build a traffic jam prediction model. The results of Random Forest classification with data attributes, namely weather, time, holidays, special conditions, and road quality produce accuracy of up to 87.5%. The advantage of this research is an effective prediction model, so that it can produce accuracy of up to 87.5%, while the disadvantage is that only a few attributes are used. Whereas in research [21] using the same classification method, namely Random Forest with 9 attributes input, namely day, date, weather, time, holiday, special conditions, starting location, destination location, and fastest route name can produce accuracy up to 92%. The advantage of this study is that the data attributes used are more, while the disadvantage of this study is that the researcher explicitly mentions that the dataset used is imbalance. In Research [12], the results of this study show that the accuracy of the CART method is 77.9%. The advantage of this research is that it uses quite a lot of attributes. But it has the disadvantage that the level of accuracy is not maximized. The difference between research [11][12][21] and this study is that in research [11][12][21] the results displayed are only in the form of accuracy while in this study the final result is a classification map.

Furthermore, in research [22], the journal discusses classification maps using the ordinary kriging method, the results in this study are regional classification maps and there is a comparison between the ordinary kriging method and other interpolation methods, namely regression kriging, the result is that the ordinary kriging method is more accurate than regression kriging with certain attribute conditions. In research [27] which discusses prediction maps using the Ordinary Kriging method with semivariogram spherical model, gaussian model, circular model, and exponential model. The result in this study is the selection of the semivariogram model is done by calculating the lowest Root Mean Square Error (RMSE) value of the four semivariogram models. Thus, in this study, ordinary kriging will be selected as an interpolation method with spherical, gaussian, and gaussian semivariogram models will be selected based on the smallest RMSE for making classification maps in this study.

Then there is research using the Google Maps application. Google Maps is a web application that can see the state of traffic congestion directly or in real time at a certain time using the Google Maps satellite [13]. In research [13] Google Maps is used as a tool to get congestion data and get Longitude and Latitude location points. So this research will also use the Google Maps application to get Longitude and Latitude location coordinates to get the location of traffic lights

Based on the above research that has been done before by considering its advantages and disadvantages, the author wants to conduct research on the topic under study with the title, "Performance Analysis of Bandung City Traffic Flow Classification with Machine Learning and Kriging Interpolation". What distinguishes the results of this study from other studies is the output produced by this study is to produce a congestion classification map in Bandung City. The expected result is that the Traffic Congestion Classification Map in Bandung City that has been made can provide solutions to traffic congestion on road sections in Bandung City.

2. RESEARCH METHODOLOGY

2.1 System Design

The system to be built is the Performance Analysis of Bandung City Traffic Flow Classification with Machine Learning and Kriging Interpolation. The following is a flowchart of the system design :

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

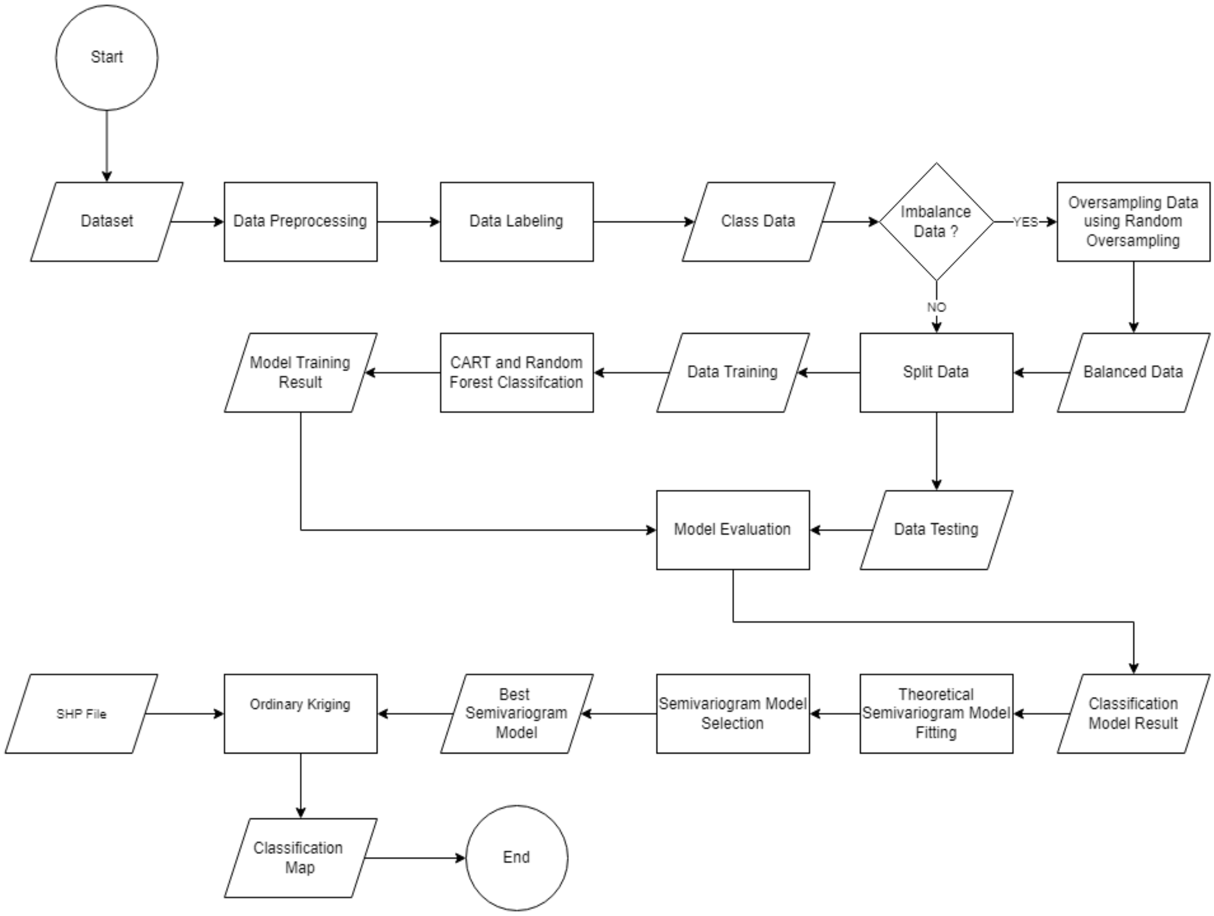

Figure 1. Flowchart Of The System Design

2.2 Dataset

The dataset used is data traffic counting from the Area Traffic Control System (ATCS) of the Bandung City Transportation Agency in April 2022. The following is an explanation of the dataset attributes and their descriptions obtained from ATCSin the following table (1) :

Table 1. Dataset from ATCS

Attribute Name	Description
Street Name	Name of Traffic Lights Points
Lane	Road Lanes
Time	Time
Motorcycle	Amount Of Motorcycle Passingby
Car	Amount Of Car Passingby
Bus/Truck	Amount Of Bus/Truck Passingby
Total	Total From Motorcycle, Car, Bus/Truck
Headway (s)	Time Interval When Vehicles Pass Through An Intersection
Gap (s)	Distance Between The Rear Of The Vehicle And The Front Of The Vehicle Behind It.
85 Speed (Km/Hour)	The Average Speed Of 85% of Vehicles On A Road Without Being Influenced By The Flow Of Traffic Whether It Is Smooth Or Congested
Avg.Speed (Km/Hour)	Average Speed Of Vehicles
Occupancy (%)	Degree Saturation Value On The Road

Based on the dataset above, additional attributes are needed to be used in making maps. To create a map, locations points are needed. To create a map, location points are needed, therefore attributes from the Google Maps web application are added, namely "Longitude" and "Latitude" to get the coordinates of “X” and “Y” of the traffic light location points. And also added new attributes for the purposes of building classifications, namely the attributes "road width" and "queue length".The final dataset can be seen in the following table (2) :

Table 2. Final Dataset

Attribute Name	Description
Street Name	Name of Traffic Lights Points

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

Longitude	Location Coordinate Point Of X
Latitude	Location Coordinate Point Of Y
Lane	Road Lanes
Time	Time
Motorcycle	Amount Of Motorcycle Passingby
Car	Amount Of Car Passingby
Bus/Truck	Amount Of Bus/Truck Passingby
Total	Total From Motorcycle, Car, Bus/Truck
Headway (s)	Time Interval When Vehicles Pass Through An Intersection
Gap (s)	Distance Between The Rear Of The Vehicle And The Front Of The Vehicle Behind It.
85 Speed (Km/Hour)	The Average Speed Of 85% of Vehicles On A Road Without Being Influenced By The Flow Of Traffic Whether It Is Smooth Or Congested
Avg.Speed (Km/Hour)	Average Speed Of Vehicles
Occupancy (%)	Degree Saturation Value On The Road
Road Width	Road Width
Queue Length	Queue Length From Traffic Light

2.3 Data Preprocessing

Datasets that are obtained and entered into dataframes can be in the form of structured or unstructured or raw data, therefore data preprocessing is needed which aims to convert these dataframes into structured dataframes for further use [10]. There are several stages in data preprocessing, namely:

a. Dataframe is Null
   This stage aims to clean the data through several processes such as checking whether there are attributes with missing or empty values.
b. Dataframe Drop Duplicate
   This stage aims to remove data that is duplicate or double.
c. Dataframe Rename
   This stage aims to change the name of the attribute column to one that is easy to understand.
d. Label Encoding
   This stage aims to convert each value in a particular column into a numeric type or consecutive numbers. Label Encoding is performed on the "Waktu" and "Hari" attributes as in the following table (3):

Table 3. Label Encoding

Attribute	Categorical Value	To Numeric Value
Time	MORNING, AFTERNOON, EVENING	0, 1, 2
Day	MONDAY, TUESDAY, WEDNESDAY, THURSDAY, FRIDAY, SATURDAY	0, 1, 2, 3, 4, 5, 6

2.4 Data Labelling

After data preprocessing is carried out, the next stage is data labeling or data labeling, data labeling is the process of providing one or more informative labels that are done manually to determine the target attribute class grouping. The target attribute class that will be processed by data labeling in this study is the "Occupancy" attribute which will be adjusted to the classification of congestion levels based on the Manual Kapasitas Jalan Indonesia (MKJI) 1997). The following is a table (4) of labeling data based on MKJI 1997:

Table 4. Data Labelling

No.	Congestion Ratio	Congestion Level	Label
1.	< 60	Free Flow	1
2.	> 60 or < 70	Stable Flow	2
3.	> 70 or < 80	Controlled Stable Flow	3
4.	≥ 80 or < 90	Unstable Flow	4
5.	> 90	Highly Congested and Delayed Flow	5

Because the "Occupancy" attribute which is used as the target class has the highest value of 86%, this research only results in the division of the congestion classification level into 4 labels or levels, namely “Free Flow”,”Stable Flow”, “Controlled Stable Flow”, “Unstable Flow”, “Highly Congested and Delayed Flow”.

2.5 Imbalance Data and Random Oversampling method

Dataframes that have been labeled and there is already a grouping of target attribute classes, will then be checked whether the target attribute class grouping has imbalance data, imbalance data is a condition where the proportion of data grouped is unbalanced or skewed [25]. To handle Imbalance Data, the Random Oversampling method will be

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

used. Random Oversampling is a method that performs the process by increasing the number of minority class samples (least) equated with the majority class samples (most) [26].

2.6 Split Data

The next stage will do data splitting on the dataframe which is divided into sub data with a splitting ratio of 80:20, 80% of the entire dataframe is used for train data or training data, and 20% of the dataframe is used for test data or testing data.

2.7 CART and Random Forest Classification

a. Classification and Regression Trees (CART)

Classification and Regression Trees (CART) is one of the decision tree algorithms that combines classification trees and regression trees. The CART algorithm starts by creating N nodes that represent the input dataframe D. If all the tuples in D have the same class value then N becomes a leaf and is labeled with the class of the tuples. To create branches, the algorithm uses a function to determine the splitting criteria. Once the best splitting criteria is found, then node N is labeled with the splitting criteria. This process repeats until there are no branches that can be created. CART has several advantages. The first advantage, this algorithm can control outliers because outliers will be separated during the splitting algorithm. Next advantage, CART will determine the important variables and eliminate the unimportant ones. Another advantage mentions that CART computation is faster[8].
CART uses a new metric named Gini Index to create decision points for classification tasks [15].
The Gini Index inequality can be expressed as follows in (1):

$$Inequality = \sum_{j} X_j \, f(r_j) \tag{1}$$

For a uniform population, the Gini Index can be expressed as follows in (2):

$$G = \frac{2\sum_{i=1}^{n} \mathrm{i} Y_i}{n\sum_{i=1}^{n} Y_i} + \frac{n+1}{n} \tag{2}$$

Gini Impurity can be expressed as follows in (3):

$$Gini\ Impurity = 1 - \sum_{i=1}^{c} P_i^2 \tag{3}$$

b. Random Forest

Random Forest is a classification method that uses several decision trees to train data samples and integrate the weights of each tree to get the output that has the most votes as the final result. The advantages of Random Forest are high classification accuracy and can handle outliers and noise in the data, therefore Random Forest is a well-known and widely used method [7][9].

Random Forest has the formula, a decision tree with M leaves divides the feature space into M regions Rm, 1 ≤ m ≤ M. For each tree, the prediction function f(x) is defined as follows in (4):

$$f(x) = \sum_{m=1}^{M} c_m \Pi\,(x, R_m) \tag{4}$$

where M is the number of a region in the feature space, Rm is a region corresponding to m; cm is a constant corresponding to m.

$$\Pi\,(x, R_m) = \begin{cases} 1, & if\ x\epsilon R_m \\ 0, & otherwise \end{cases} \tag{5}$$

Then the final classification conclusion is made from the most votes as the final result of all trees.

2.8 Model Evaluation

In the Model Evaluation stage process aims to determine the performance of each classification model method used. At the Model Evaluation stage using the confusion matrix method, the confusion matrix is used to measure the performance of the classification method and uses a table with 4 combinations of predicted values and actual values shown in table 6. In this study using confusion matrix with several statistical measures such as accuracy, precision, recall, and f1-score as a performance measurement tool [17][21].

Table 5. Confusion Matrix

Predicted Values	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

With information, TP is True Positive, FP is False Positive, TN is True Negative, FN is False Negative.

Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

Based on the confusion matrix, the performance of the classification method can be calculated with statistical measures such as the following classifications built among others:

a. Accuracy

Accuracy is the ratio value of the number of correct predictions to the total number of all classified predictions [13]. Accuracy can be formulated as follows (6) :

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (6)$$

b. Precision

Precision is the ratio value that predicts true positives compared to the overall positive predicted results [13]. Precision can be formulated as follows (7) :

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (7)$$

c. Recall

Recall or also called sensitivity is the ratio value of true positive predictions compared to the overall true positive data [13]. Recall can be formulated as follows (8) :

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (8)$$

d. F1-Score

F1-Score is a comparison of weighted average precision and recall [13]. F-Score can be formulated as follows (9) :

$$\text{F1} - \text{Score} = \frac{2(\text{PrecisionxRecall})}{\text{Precision+Recall}} \quad (9)$$

2.9 Semivariogram Model Selection

At the semivariogram model selection stage, the best semivariogram model selected is a semivariogram model that has been compared to the Root Mean Square Error (RMSE) value of the three semivariogram models, namely the spherical model, exponential, and gaussian model. RMSE is an alternative method for evaluating prediction cases by measuring the error rate of the prediction results. The general form of RMSE is formulated as follows (10) :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \quad (10)$$

Where Y is the actual value of the test data, Ŷ is the prediction value of the test data, and finally n is the number of test data.

The semivariogram model that has been selected will then be input as a model in the interpolation process using the Ordinary Kriging method.

2.10 Ordinary Kriging

Ordinary Kriging is one of the most commonly used kriging techniques. It predicts the spatial unmeasured at a given pointx_0 by predicting the value of$Z^*(x_0)$, which is equal to the number of known measurement lines [22]. Here is a simple Ordinary Kriging formula (11) :

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad (11)$$

Where $Z^*(x_0)$ is the predicted value at the unmeasured position x_0, $Z(x_i)$is the measured value at position $x_i\lambda_i$, is the weighting coefficient from the measured position to x_0 and n is the number of positions in the neighborhood search [22]. The best semivariogram model will be selected and input to the Ordinary Kriging method. In this research, the Ordinary Kriging method is calculated using ArcMap software in the Geostatistical Analyst method section.

2.11 Classification Map

Classification Map is the process of giving color to the map of areas in Bandung City based on the results of the classification of congestion levels from the CART and Random Forest methods that have been obtained [18]. With the first stage carried out, the data from the results of the Classification And Regression Trees (CART) and Random Forest models obtained will then be added to the ArcMap software or software. Then Shapefiles (SHP) of roads in Bandung City are inputted as an initial map of location observations, so that location data can be read by ArcMap. After inputting data from the results of the Classification And Regression Trees (CART) method model, Random Forest and the Bandung City SHP file, the next step is to interpolate using Ordinary Kriging with the semivariogram of the selected model so as to produce a congestion classification map. The resulting congestion classification map will be a map of the area with roads colored according to the results of Ordinary Kriging which shows each road with a certain color according to the results obtained.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

3. RESULT AND DISCUSSION

In this study, two test scenarios will be carried out, the first test conducted is comparing the accuracy of the two CART and Random Forest methods on Imbalance Datasets and Datasets that have been Random Oversampling. Furthermore, the second test will create two classification maps of traffic congestion in Bandung City and analyze the two classification maps that have been made.

3.1 Dataset

This research uses datasets that have passed the data preprocessing process, data labeling, and produce target class data in accordance with the classification of congestion levels based on MKJI 1997. The dataset in this study is 3804 which is divided into 4 labels, namely “Free Flow”, ”Stable Flow”, “Controlled Stable Flow”, dan “Unstable Flow” can be seen in Figure 2 :

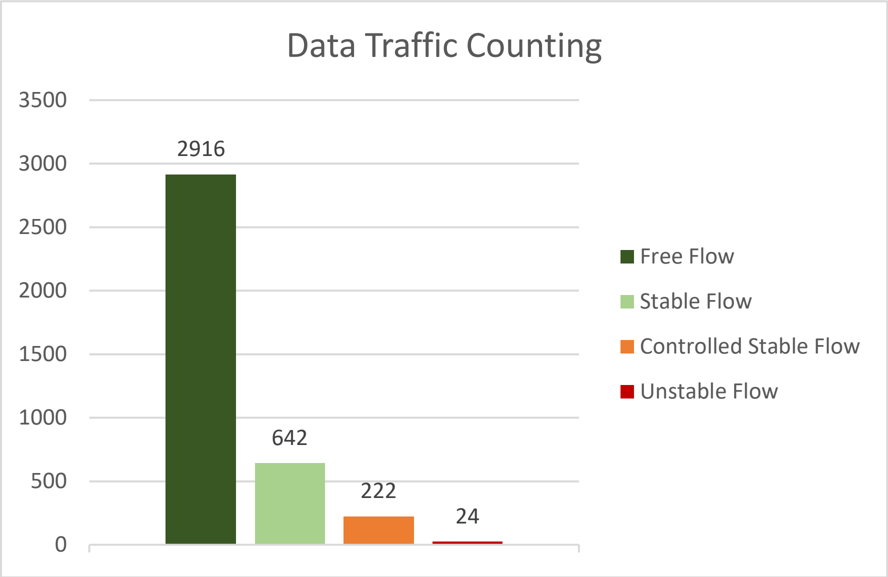

Figure 2. Data Traffic Counting

Furthermore, data splitting will be carried out on the dataset with a splitting ratio of 80: 20, with the information that 80% of the dataset will be used to train data or training data and 20% to test the model built shown in Figure 3.

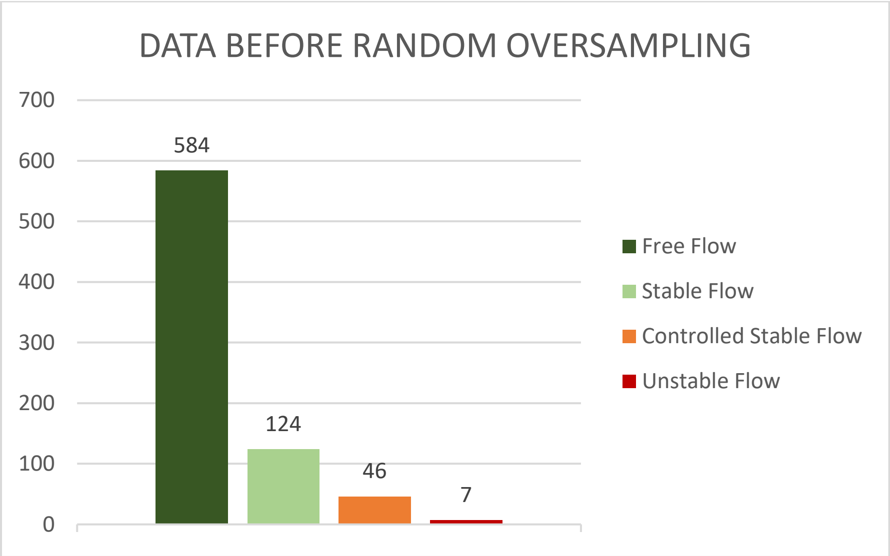

Figure 3. Imbalance Data

In Figure 3 shows the occurrence of Data Imbalance in the test data, Imbalanced Data is a condition where the proportion of data grouped is unbalanced or skewed which will result in learning classification models will tend to be smarter to predict data classes with more data such as in the “Free Flow” data will be much smarter. To handle Imbalance Data, the Random Oversampling method will be used. Random Oversampling is a method that performs the process by increasing the number of minority class samples (the least) equalized with the majority class samples (the most). Here is the data after Random Oversampling in Figure 4.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

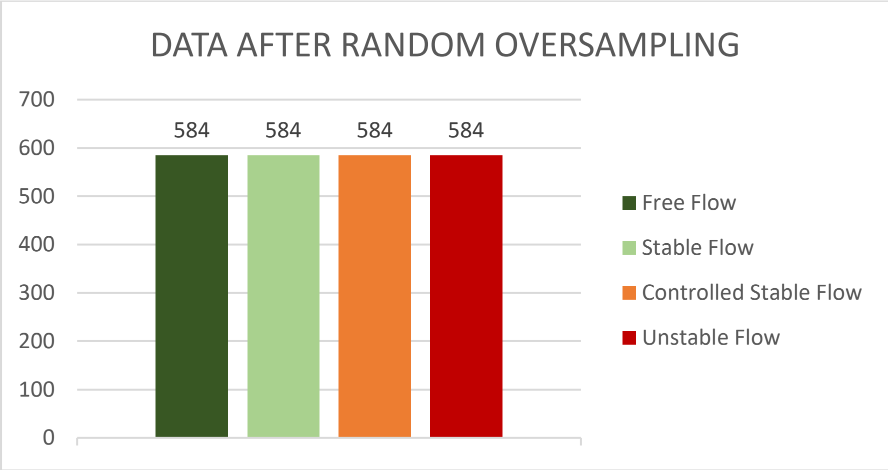

Figure 4. Imbalance Data

In Figure 4, it can be seen that all data classes increase with the same proportion of data of 584 for all four labels, this is because Random Oversampling increases the number of samples or the minority class data is equalized with the majority class samples (at most).

3.2 Model Evaluation

At the Model Evaluation stage, a comparison of the performance of the two classification models, namely, Classification And Regression Tree (CART) and Random Forest, will be carried out. Two (2) test scenarios will be carried out with the first test, namely calculating the performance of the two classification models with data imbalance without using Random Oversampling can be seen in Table 6.

Table 6. Model Evaluation with Imbalance Data

Model	Accuracy	Precision	Recall	F1-Score
CART	0.99	0.99	0.99	0.99
Random Forest	0.99	0.99	0.99	0.99

Results in the first test with data imbalance without using Random Oversampling. Shows very high accuracy results with an accuracy rate of 99% for both CART and Random Forest models. With a record for the precision, recall and f1-score values for both methods getting 99% for each model. This happens because the data used to build the model is imbalance, so the model is only smart to classify the majority class data, not the minority class data. Next, the second test will be carried out, namely calculating the performance of the two classification models using Random Oversampling, which can be seen in Table 7.

Table 7. Model Evaluation using Random Oversampling

Model	Accuracy	Precision	Recall	F1-Score
CART	0.88	0.91	0.88	0.87
Random Forest	**0.90**	**0.91**	**0.90**	**0.90**

The results in the second test using Random Oversampling show that the Random Forest model is superior to the CART model. With the information that the Random Forest model gets an accuracy level of 90%, precision 91%, recall 90%, and f1-score 90%. CART gets an accuracy level of 88%, precision 91%, recall 88%, and f1-score 87%. This proves that the Random Forest model is superior and can be good at building classification models using Random Oversampling.

3.3 Semivariogram Selection

Table 8. Theoretical semivariogram calculation results with parameters: nugget, major range, and minor range.

Method	Semivariogram	Nugget	Major Range	Minor Range
CART	Exponential	1,04736	0,1343156	0,05481813
	Gaussian	**0,9743184**	**0,1343156**	**0,04511438**
	Spherical	1,001889	0,1343156	0,04736606
Random Forest	Exponential	1,181918	0,1311317	0,06698455
	Gaussian	**1,098877**	**0,1311317**	**0,0485592**
	Spherical	1,131435	0,1311317	0,05188405

Table 9. RMSE and theoretical semivariogram calculation results with parameters: direction and partial sill

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

Metode	Semivariogram	Direction	Partial Sill	RMSE
CART	Exponential	92,46094	0,1110785	0,90101
	Gaussian	**99,84375**	**0,2943507**	**0,89820**
	Spherical	95,97656	0,1936963	0,89945
Random Forest	Exponential	91,23047	0,1148437	0,94330
	Gaussian	**100,1953**	**0,3269722**	**0,93631**
	Spherical	96,15234	0,20781106	0,94062

The results of theoretical and RMSE calculations in tables 8 and 9 show that theoretical semivariograms with nugget, major range, minor range, direction, and partial sill parameters can produce RMSE for exponential model, gaussian model, and spherical model semivariograms for each CART and Random Forest method. Furthermore, the semivariogram with the **smallest** RMSE for each CART and Random Forest method is selected, so the gaussian semivariogram model for the CART method is selected with an RMSE of **0.89820** and the gaussian semivariogram model for the Random Forest method is also selected with an RMSE of **0.93631**. Therefore, the Gaussian semivariogram model for both CART and Random Forest classification methods will be selected for the interpolation process in Ordinary Kriging to create a classification map of traffic congestion in Bandung City.

3.4 Classification Map

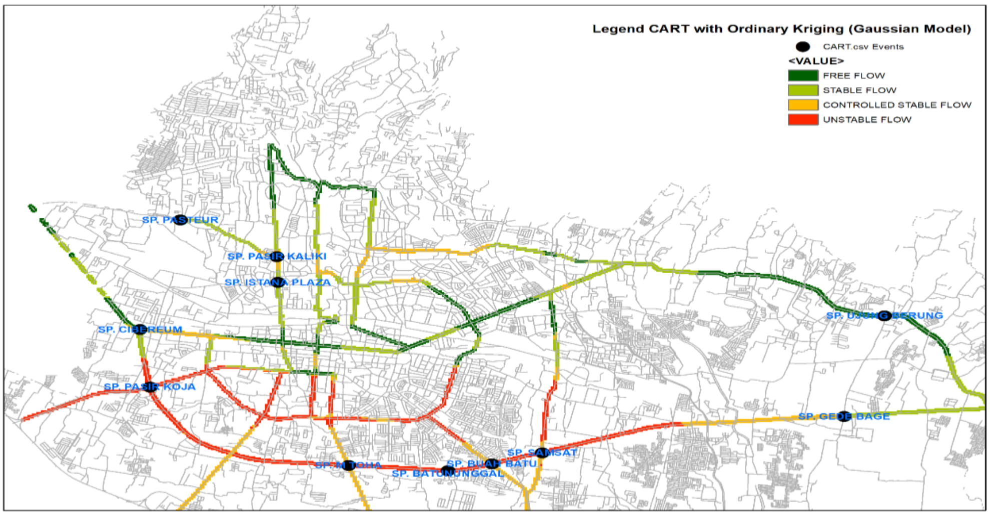

Figure 5. Classification map using CART method with Ordinary Kriging (Gaussian Model)

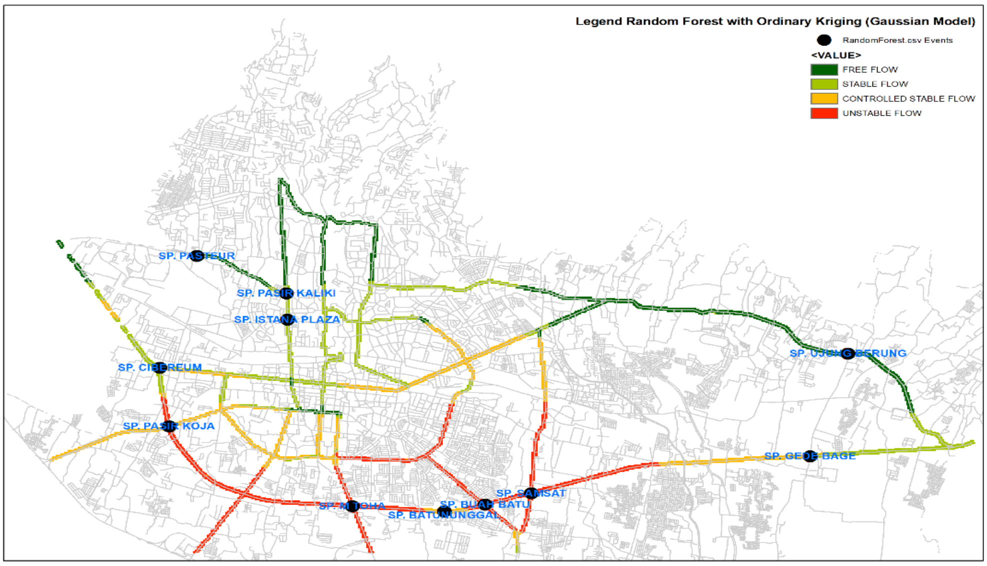

Figure 6. Classification map using Random Forest method with Ordinary Kriging (Gaussian Model)

Traffic congestion classification map in Bandung City using CART and Random Forest methods with Ordinary Kriging interpolation (Gaussian Model) is shown in Figures 5 and 6. The interpolation results are displayed visually in the form of maps using ArcMap software. The distribution of congestion levels is divided into four categories based on the "Occupancy" attribute and according to the 1997 Indonesian Road Capacity Manual (MKJI) produces 4

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

categories, namely: "Free Flow" marked in dark green, "Stable Flow" marked in light green, "Controlled Stable Flow" marked in orange, and "Unstable Flow" marked in red. Each SP or stop point is indicated by a black round dot.

The results of the analysis of the level of congestion in Bandung City show that the northern part of Bandung City tends to be stable by having differences in both the CART and Random Forest methods, the CART method shows that at the point SP. PASTEUR both methods show different results with the CART method having "Stable Flow" which is shown in light green, while in the Random Forest method the point SP. PASTEUR has a "Free Flow" which is shown in dark green. Furthermore, the CART method at point SP. PASIR KALIKI with a connection to point SP. ISTANA PLAZA has "Controlled Stable Flow" and "Stable Flow" which are shown in orange and light green. While in the Random Forest method at the point SP. PASIR KALIKI with a connection to the point SP. ISTANA PLAZA has "Free Flow" and "Stable Flow" which are shown in dark green and light green. In the western part of Bandung City shows the difference in results from both CART and Random Forest methods in the form of the CART method point SP. CIBEREUM has "Free Flow" and "Stable Flow" indicated by the presence of dark green and light green colors, while in the Random Forest method point SP. CIBEREUM has only "Stable Flow". But in the western part of Bandung City at point SP. PASIR KOJA both CART and Random Forest methods have the same flow, namely "Unstable Flow" shown in red. Furthermore, in the southern part of Bandung City, it tends not to be smooth with the currents owned, namely "Controlled Stable Flow" and "Unstable Flow" with slight differences in the two methods. The difference is that in the CART method at the point SP. M TOHA, SP. BATU NUNGGAL, SP. BUAH BATU, and SP. SAMSAT shows the presence of "Unstable Flow" indicated in red, while in the Random Forest method the points at SP. M TOHA, SP. BUAH BATU, and SP. SAMSAT shows that it only has one "Unstable Flow" indicated in red, but at the point SP. BATU NUNGGAL shows a difference in the Random Forest method at point SP. BATU NUNGGAL has "Unstable Flow" shown in red. Turning to the eastern part of Bandung City, both methods show the same results, at point SP. UJUNG BERUNG has "Free Flow" shown in dark green and SP. GEDE BAGE has "Stable Flow" shown in light green with a continuation from point SP. GEDE BAGE to SP. SAMSAT has "Controlled Stable Flow" shown in orange.

3.5 Discussion

With the results of the classification map that has been made, in the northern area of Bandung City, the proportion of the level of congestion has a stable flow towards controlled stable flow. Then in the western part of the proportion of congestion has two different flows, namely stable flow and unstable flow. While in the eastern part has two flows, namely free flow and stable flow. Finally, the southern section is a road with a consistent level of congestion that only has an unstable flow.

Comparison of this study with other similar studies, in research [11][21] also used the Random Forest method, with the accuracy results in research [11] resulting in 87.5% accuracy and in research [21] resulting in 92% accuracy, while in this study the results of the Random Forest method obtained an accuracy of up to 90%, this proves that the Random Forest method can build classification models with a fairly high level of accuracy.Furthermore, Research [12] using the CART method also produces accuracy of up to 77.9%, while in this study the CART method can produce accuracy of up to 88%, this also proves that the CART method can produce quite high accuracy.Finally, the same research [23] also created a classification map but with different output results from the type of classification map with this research. In [23], classification maps were created with many types of maps such as area classification maps, road classification maps, and regional classification maps. Whereas in this research only the results of the type of road classification map will be made.

Therefore, by looking at the results of the classification map that has been made, the author wants to provide two solutions to traffic congestion in Bandung City, especially in the southern part of Bandung City which has Unstable Flow, the solutions provided are, the first solution is to reduce the use of private vehicles and use more public transportation. The second solution is the implementation of "Open and Close Road System", "Open and Close Road System" can be done at certain hours with certain flow conditions. If on a road with a stable flow, a two-way road system will be applied, while if on a road that shows an unstable flow, a one-way road system will be applied. It is expected that with the solutions that have been given, the congestion that has been happening can be reduced.

4. CONCLUSION

The results of the comparison of classification maps made with Ordinary Kriging interpolation with the Gaussian Model semivariogram in both methods, namely CART and Random Forest. With the CART method has an accuracy of up to **88%** while the classification map made with the Random Forest method has an accuracy of up to **90%.** This proves that in this study the **Random Forest method is far superior** in building classification maps compared to the CART method. So it can be concluded that from the results of the classification map of traffic congestion in Bandung City that has been made can produce a good classification map with the help of Random Oversampling to overcome Data Imbalance. With the results of the classification map that has been made, it can be concluded with the division of congestion in the northern, western, eastern, and southern parts of Bandung City. In the northern part of Bandung City, the proportion of congestion levels has a stable flow towards controlled stable flow. Then in the western part, the proportion of congestion levels has two different flows, namely stable flow and unstable flow. While in the eastern

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 694−704
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1972

part has two flows, namely free flow and stable flow. Finally, the southern section is a road with a consistent level of congestion that only has an unstable flow.

REFERENCES

[1] BPS Kota Bandung, "Jumlah Penduduk Kota Bandung," *Jumlah Penduduk Kota Bandung*, 2020. https://bandungkota.bps.go.id/linkTableDinamis/view/id/9 (accessed Nov. 19, 2021).
[2] E. Harahap, A. Suryadi, R. Ridwan, D. Darmawan, and R. Ceha, "Efektifitas Load Balancing Dalam Mengatasi Kemacetan Lalu Lintas," *Matematika*, vol. 16, no. 2, pp. 1–7, 2017, doi: 10.29313/jmtm.v16i2.3665.
[3] V. Effendy, "Sentiment Analysis on Twitter about the Use of City Public Transportation Using Support Vector Machine Method," *Int. J. Inf. Commun. Technol.*, vol. 2, no. 1, p. 57, 2016, doi: 10.21108/ijoict.2016.21.85.
[4] M. Somantri, "Kemacetan dan Pandemi di Kota Bandung (24 Juni 2021)," 2021. https://kumparan.com/musliman-somantri-1614653043992639605/kemacetan-dan-pandemi-di-kota-bandung-1w0JT6DMtQq/full (accessed Nov. 19, 2021).
[5] B. N. Mohapatra and P. P. Panda, "Machine learning applications to smart city," *Accent. Trans. Image Process. Comput. Vis.*, vol. 5, no. 14, pp. 1–6, 2019, doi: 10.19101/tipcv.2018.412004.
[6] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons.B*, vol. 4, no. December 2017, pp. 51–62, 2017, doi: 10.20544/horizons.b.04.1.17.p05.
[7] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00327-4.
[8] R. Irmanita, Sri Suryani Prasetiyowati, and Yuliant Sibaroni, "Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 10–16, 2021, doi: 10.29207/resti.v5i1.2770.
[9] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.
[10] K. Nugroho, "Dasar Text Preprocessing dengan Python," 2019. https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe (accessed Dec. 05, 2021).
[11] Y. Liu and H. Wu, "Prediction of road traffic congestion based on random forest," *Proc. - 2017 10th Int. Symp. Comput. Intell. Des. Isc. 2017*, vol. 2, pp. 361–364, 2018, doi: 10.1109/ISCID.2017.216.
[12] L. Susiana, I. T. Utami, and J. Junaidi, "Penerapan Metode Boosting Pada Cart Untuk Mengklasifikasikan Korban Kecelakaan Lalu Lintas Di Kota Palu," *Nat. Sci. J. Sci. Technol.*, vol. 8, no. 2, pp. 106–109, 2019, doi: 10.22487/25411969.2019.v8.i2.13536.
[13] N. Petrovska and A. Stevanovic, "Traffic Congestion Analysis Visualisation Tool," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2015-Octob, pp. 1489–1494, 2015, doi: 10.1109/ITSC.2015.243.
[14] S. Ki, M. Keffeler, T. Atkison, and A. Hainen, "Using Time Series Forecasting for Adaptive Traffic Signal Control," *Proc. 13th Int. Conf. Data Min. DMIN'17*, pp. 34–39, 2017.
[15] T. Daniya, M. Geetha, and K. S. Kumar, "Classification and regression trees with gini index," *Adv. Math. Sci. J.*, vol. 9, no. 10, pp. 8237–8247, 2020, doi: 10.37418/amsj.9.10.53.
[16] M. I. Ali and M. R. Abidin, "Pengaruh Kepadatan Penduduk Terhadap Intensitas Kemacetan Lalu Lintas Di Kecamatan Rappocini Makassar," *Pros. Semin. Nas. Lemb. Penelit. Univ. Negeri Makassar*, pp. 68–73, 2019, [Online]. Available: http://eprints.unm.ac.id/17795/1/prosiding Pengaruh Kepadatan Penduduk.pdf
[17] T. S. Tamir *et al.*, "Traffic Congestion Prediction using Decision Tree, Logistic Regression and Neural Networks," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 512–517, 2020, doi: 10.1016/j.ifacol.2021.04.138.
[18] N. Saif and M. Mussafi, "Penerapan Greedy Coloring Algorithm Pada Peta Kotamadya Yogyakarta Berbasis Four-Colour Theorem," *Kaunia Integr. Interconnect. Islam Sci.*, vol. 11, no. 1, pp. 19–26, 2015, [Online]. Available: http://ejournal.uin-suka.ac.id/saintek/kaunia/article/view/1078
[19] L. E. Edison, "Analisis Dampak Kerugian Akibat Kemacetan Lalu Lintas Di Kota Makassar.," 2017.
[20] A. Aid, M. A. Khan, S. Abbas, G. Ahmad, and A. Fatimat, "Modelling smart road traffic congestion control system using machine learning techniques," *Neural Netw. World*, vol. 29, no. 2, pp. 99–110, 2019, doi: 10.14311/NNW.2019.29.008.
[21] N. Zafar and I. U. Haq, "Traffic congestion prediction based on Estimated Time of Arrival," *PLoS One*, vol. 15, no. 12 December, pp. 1–19, 2020, doi: 10.1371/journal.pone.0238200.
[22] T. G. Pham, M. Kappas, C. Van Huynh, and L. H. K. Nguyen, "Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central Vietnam," *ISPRS Int. J. Geo-Information*, vol. 8, no. 3, 2019, doi: 10.3390/ijgi8030147.
[23] J. Song, C. Zhao, S. Zhong, T. A. S. Nielsen, and A. V. Prishchepov, "Mapping spatio-temporal patterns and detecting the factors of traffic congestion with multi-source data fusion and mining techniques," *Comput. Environ. Urban Syst.*, vol. 77, no. July, p. 101364, 2019, doi: 10.1016/j.compenvurbsys.2019.101364.
[24] H. Hardiani, "Analisis Derajat Kejenuhan dan Biaya Kemacetan pada Ruas Jalan Utama di Kota Jambi," *J. Perspekt. Pembiayaan dan Pembang. Drh.*, vol. 2, no. 4, pp. 181–192, 2016, doi: 10.22437/ppd.v2i4.2614.
[25] J. Gong and H. Kim, "RHSBoost: Improving classification performance in imbalance data," *Comput. Stat. Data Anal.*, vol. 111, no. xxxx, pp. 1–13, 2017, doi: 10.1016/j.csda.2017.01.005.
[26] 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.
[27] S. S. Prasetiyowati, M. Imrona, I. Ummah, and Y. Sibaroni, "Prediction of public transportation occupation based on several crowd spots using ordinary kriging method," *J. Innov. Technol. Educ.*, vol. 3, no. January, pp. 93–104, 2016, doi: 10.12988/jite.2016.6723.