E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

BANDUNG CITY TRAFFIC CLASSIFICATION MAP WITH MACHINE LEARNING AND ORDINARY KRIGING

Winico Fazry[1], Sri Suryani Prasetyowati[2], and Yuliant Sibaroni[3]
[1, 2, 3] Informatics, School of Computing, Telkom University
Jl. Telekomunikasi No 1, Bandung
e-mail: winicofazry@student.telkomuniversity.ac.id[1], srisuryani@telkomuniversity.ac.id[2], yuliant@telkomuniversity.ac.id[3]

ABSTRAK

Kemacetan merupakan masalah yang terjadi ketika jumlah kendaraan melebihi kapasitas jalan, dan kecepatan kendaraan melambat. Isu ini merupakan salah satu isu utama di kota-kota besar, termasuk Bandung. Dalam penelitian ini, penelitian ini bertujuan untuk mengurangi kemacetan lalu lintas di Kota Bandung. Proses klasifikasi pada penelitian ini menggunakan metode Support Vector Machine (SVM), Naive Bayes, dan Ordinary Kriging. Data yang digunakan adalah data traffic counting dari ATCS kota Bandung dan observasi langsung. Data penghitungan lalu lintas yang diperoleh berisi 3804 baris. Tiga skenario eksperimen dilakukan untuk memvalidasi efektivitas model yang digunakan, kinerja model pertama tanpa oversampling, kinerja model kedua dengan oversampling, dan kinerja model ketiga dengan penyetelan hyperparameter. Hasil eksperimen menunjukkan bahwa metode Support Vector Machine memiliki akurasi yang lebih tinggi dibandingkan dengan metode Naive Bayes yaitu sebesar 93%, sedangkan metode Naive Bayes memiliki akurasi sebesar 90%. Penerapan hyperperparameter tuning dan over-sampling terbukti dapat mengatasi masalah ketidakseimbangan data dan mendapatkan hasil klasifikasi yang lebih baik. Selain itu, hasil klasifikasi terbaik digunakan dalam pembuatan peta klasifikasi yaitu metode Support Vector Machine, dan dibantu dengan kriging biasa untuk memprediksi daerah sekitar. Hasil peta klasifikasi kemacetan menunjukan daerah selatan pada kota Bandung menunjukan lebih tidak stabil dibandingkan daerah kota bandung lainya.

***Kata Kunci**: Kemacetan, Klasifikasi, Naïve Bayes, Ordinary Kriging, Support Vector Machine.*

ABSTRACT

Congestion is a problem that occurs when the number of vehicles exceeds the capacity of the road and the vehicle speed slows down. This issue is one of the main issues in big cities, including Bandung. In this study, this study aims to reduce traffic congestion in the city of Bandung. The classification process in this study uses the Support Vector Machine (SVM), Naive Bayes, and Ordinary Kriging methods. The data used is traffic counting data from ATCS in Bandung and direct observation. The traffic count data obtained contains 3804 rows. Three experimental scenarios were carried out to validate the effectiveness of the model used, the performance of the first model without oversampling, the performance of the second model with oversampling, and the performance of the third model with hyperparameter adjustment. The experimental results show that the Support Vector Machine method has higher accuracy than the Naive Bayes method, which is 93%, while the Naive Bayes method has an accuracy of 90%. The application of hyperparameter tuning and over-sampling is proven to overcome the problem of data imbalance and get better classification results. In addition, the best classification results are used in making classification maps, namely the Support Vector Machine method, and assisted with ordinary kriging to predict the surrounding area. The results of the congestion classification map show that the southern area of the city of Bandung is more unstable than other areas of the city of Bandung.

***Keywords**: Congestion, Classification, Naïve Bayes, Ordinary Kriging, SVM.*

I. INTRODUCTION

TRAFFIC congestion is a widespread global phenomenon due to high population density, the growth of motor vehicles and infrastructure, and the proliferation of transportation and delivery services[1]. The researchers identified the barriers from different angles. The most common definition of congestion in traffic conditions is when travel demand exceeds road capacity[2]. The issue of congestion has become one of the main problems in big cities in Indonesia, including in the city of Bandung, the city of Bandung has also become one of the cities that become tourist destinations for domestic and international tourists, which is the impact of the increasing number of vehicles entering this city causing heavy traffic jams Occurred in the town of Bandung[3]. Many factors cause congestion in the city of Bandung, such as the number of vehicles, public facilities, the number of red lights, and the width of the road. Due to these factors, the traffic flow of road users is crowded on the highway[4]. By looking at the factors that cause congestion in the city of Bandung, a solution is needed to solve the factors that cause congestion in the city of Bandung. Along with the development of science and technology and natural science,

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

machine learning has attracted much attention[5]. The story of machine learning is proliferating and has overcome many existing problems. And this machine learning can also predict roads that are experiencing congestion and can make it easier to make a congestion prediction map so that it can make it easier to choose routes that do not experience congestion[6].

Research [6] uses the Support Vector Machine method to detect vehicle congestion in Bandung. This study uses precision, recall, and accuracy as test parameters to detect vehicle congestion in Bandung. The results obtained from the test have an accuracy rate of 89.62%, a precision of 85.68%, and a recall of 96.62%, which shows that the use of the Support Vector Machine method has a high level of accuracy. However, this research lacks in processing a longer path and further research [7], still using the Support Vector Machine method to predict traffic flow in Minneapolis. This study selected five denoising models, namely EMD, EEMD, MA, BW, and WL, to evaluate the effect of predictive performance. The results obtained by EEMD can produce the highest prediction results, the prediction accuracy of EMD and WL are both still slower than EEMD, and the performance of MA is the lowest.

Another study [8] used the Naïve Bayes method and SVM to detect traffic congestion in Indonesia, especially in DKI Jakarta. In this study, it was found that the SVM method was higher than the Naïve Bayes method in terms of prediction accuracy. The Naïve Bayes method obtained an accuracy rate of 86.62%, while the SVM method obtained an accuracy of 96.29%. Further research [9] uses Naïve Bayes to predict traffic flow. In this study, the Naïve Bayes method obtained an accuracy of 82.7%, but only time and area factors were considered in predicting traffic flow. The following study [10] uses ordinary kriging to explore the randomness of the pattern of data lost in traffic and improve the prediction results by other classification methods. Subsequent research [11] uses ordinary kriging to help improve short-term traffic volume predictions because the classification results using other classification methods are still lacking in predicting short-term traffic volumes. Therefore this study uses ordinary kriging to help improve prediction results.

Based on the results of previous studies encourages the author to examine the classification of traffic congestion maps in the city of Bandung. Because from previous research, no one has made a traffic congestion classification map, only classifying congestion. The author uses a classification method using a support vector machine (SVM) and Naïve Bayes because it has high accuracy in traffic congestion classification and is also coupled with ordinary kriging to explore the randomness of data patterns lost in traffic. To get the highest accuracy results in predicting congestion and congestion classification maps can reduce traffic jams in the city of Bandung.

II. Research Methodology

A. system design

The system to be built is traffic congestion classification using the Support Vector Machine (SVM) and Naïve Bayes methods. The following is a flowchart of the system design that was built:

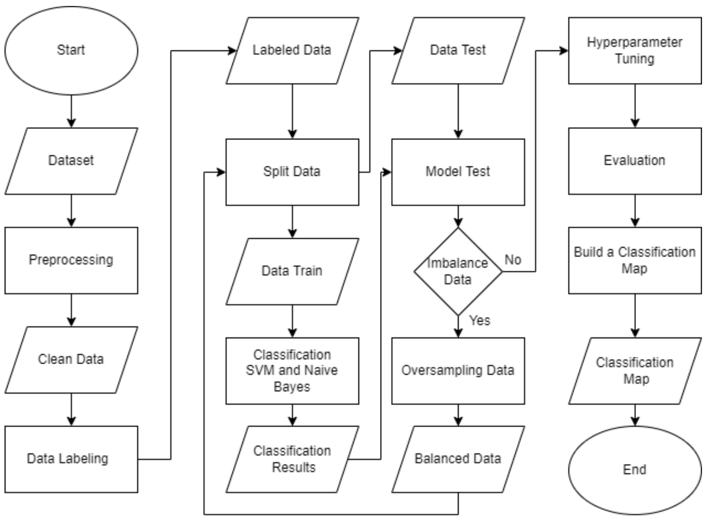

Fig. 1. Flowchart

Based on Fig 1 will be explained the flow of the system.

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

B. Dataset

This study uses traffic counting data from ATCS in the city of Bandung and also direct observation. Traffic counting data contains 3804 rows from April 1 - April 30, 2022. Traffic counting data contains street names, lanes, time, day, date, motorbike, car, bus/truck, total, headway, GAP, 85p speed, AVG. Speed, occupancy, and additional data from observations containing columns of latitude, longitude, road width, and road length. Tables I and II below show examples of the dataset used.

TABLE I
DATASET

Street Names	Latitude	Longitude	Lanes	Time	Day	Date	Motor-bike	Car
SP. BUAH BATU	6.947987	107.633434	Straight Right	Afternoon	Thursday	07-04-2022 16:00	479	479
SP. SAMSAT	6.945438	107.641889	Straight Right	Morning	Tuesday	12-04-2022 7:00	4232	4232
SP. BUAH BATU	6.947987	107.633434	Straight Right	Morning	Thursday	21-04-2022 07:00	1202	1202

TABLE II
DATASET

Bus/Truck	Total	Headway	GAP	85p Speed	AVG. Speed	Occupancy	Road Width	Queue Length
53	824	0.37	10.28	47.50	21.31	86.30	21.0	700
139	4577	0.00	1.71	60.25	11.94	85.21	21.0	700
107	2055	0.00	2.36	39.25	12.62	85.21	21.0	700

C. Preprocessing

Preprocessing is an important initial stage to reduce or modify raw data or noise amounts before classification because performing direct analysis on dialect texts can lead to poor results[12]. This study uses preprocessing split data 70% training data and 30% test data and also labeling data on attributes that have categorical values into numerical values and labeling data on congestion level classification based on Indonesian Road Capacity Manual 1997[13][14].

D. Labeling Data

Accurate data labeling from raw data is needed for the next step, and data labeling is essential so that previous data can improve traffic prediction performance and congestion classification. An example can be seen in Table III, converting categorical values to numeric[15].

TABLE III
LABELING DATA

Attribute	Categorical Value	Numerical Value
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday	0, 1, 2, 3, 4, 5, 6
Time	Morning, afternoon. Evening	0, 1, 2

Table IV. Labeling occupancy data into congestion levels based on the Indonesian Road Capacity Manual (1997), the levels are divided into 4, namely level 0 is free to flow, level 1 is stable flow, level 2 is stable and controlled, and level 3 is unstable.

TABLE IV
CONGESTION LEVEL LABELING

Attribute	Attribute value	Congestion Rate	Lebel
Occupancy	$\leq 60\%$	Free Flow	0
	$>60\% \leq 70\%$	Steady flow	1
	$>70\% \leq 80\%$	Stable Controlled	2
	$>80\% \leq 90\%$	Unstable	3

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

E. Data Oversampling

Data oversampling is needed if the data obtained are not balanced. Balanced data is a state of data that is not balanced between data classes 1 in other data classes. The disproportionate state of the data is a classification problem because classifier training tends to predict many (mostly) and small (few) classes. This provides predictive accuracy suitable for training data (majority class) across multiple classes. If the train data (minority class) is small, the prediction accuracy will decrease[16]. Due to the significant imbalance in this dataset, data oversampling is needed. The data oversampling that the author uses is random oversampling to improve prediction accuracy[17].

F. Implementation

1. *Support Vector Machine*

Support Vector Machine (SVM) is still one of the classic machine learning methods to help solve big data classification problems. It can also help multi-domain applications in big data environments. However, vector machine support is mathematically complex and computationally expensive[18]. Support Vector Machine (SVM) is a classification method technique that uses machine learning (supervised learning) to predict classes based on models or patterns obtained from the training process[19]. The linear function used in SVM can be expressed as [20]:

$$f(x) = (\omega, x) + b \tag{1}$$

Where ω represents the weight value as the coefficient for each feature. This value is adjusted during processing. The algorithm tries to find the weight of the value used to create the hyperplane with the most significant margin. There are several kernel functions commonly used in SVM, such as RBF, polynomial and sigmoid, which are stated as follows[20]:

a. Radial basis Function (RBF)

$$K(x_i, x_j) = \exp\left(-y \parallel x_i, x_j \parallel\right)^2 \tag{2}$$

b. Polynomial

$$K(x_i, x_j\ ) = \left(yX_i^T\ x_j + r\right)^p, y > 0 \tag{3}$$

c. Sigmoid

$$K(x_i, x_j) = tanh\ (x_i^T x_j + r) \tag{4}$$

2. Naïve *Bayes*

A naive Bayes classifier is a particular Bayesian network designed for problem classification. This Naive Bayes classification is a simple probabilistic classification model that calculates the probability of a given target variable or class of variables. For example, feature or attribute variables and then predict the class of the target variable with high probability. Compared to other machine learning models, the computational simplicity of the Naive Bayes method and the Naive Bayes method can be trained very quickly. Naive Bayes methods usually contain one target variable and several feature variables. Suppose T is the state or class of the target variable, and the vector X = (x1,x2,...,xn) becomes the state of n features. Simply the value of T is based on X. The following is the calculation process[21]:

$$p(T|X) = \frac{p(X|T)p(T)}{(p(X))} \tag{5}$$

Where p(X) and p(T) are constants that can be derived directly from the data, while p(X|T) is left to solve. Based on the assumption of Naïve Bayes feature independence, p(X|T) can be factored into[21]:

$$p(X|T) = p(x_1, x_{2,}, \dots, x_n|T) = \prod_{i=1}^{n}(x_i\ \ |T) \tag{6}$$

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

By combining the two previous equations, we have [21]:

$$p(T \mid X) = \frac{p(t)}{(p(x))} \prod_{i=1}^{n} p(x\_i \ |T) \tag{7}$$

where p(T), p(X), and p(Xi|T) are the parameters of the Naïve Bayes model, the parameters are obtained from the training data. The distribution of T and given X can be used in the equation.

3. Ordinary *Kriging*

Ordinary Kriging is said to be the easiest method of geostatistical data. This method assumes that the population mean (mean) is not. It is known but has a constant value, and the spatial data used does not include trends and outliers [22]. Estimating {y(x)∈ A} denotes the spatial response variable of position x. It is known that the value of an attribute at each sample point X = {xi, i = 1, 2, . . . , n} is y(xi), the estimated value of y(x0) at the unsampled position x0 is the weighted sum of value X [23]:

$$\hat{y}(x_0) = \sum_{i=1}^{n} \lambda_i y(x_i) \tag{8}$$

Whare ʎi The appropriate weight parameters for each sample point are determined by the variogram, which defines the variance of the difference between values at two locations, much like the exponential model.

$$\gamma(h) = c_0 + C(1 - e^{-\frac{h}{a}}) \tag{9}$$

Where h is the distance between the two sample points, a is the range, C0 represents the nugget effect, and C represents the partial threshold. The last three parameters were fitted from the empiric variogram:

$$\hat{y}\ (h + \vartheta h) = \frac{1}{2N\ (h + \vartheta h)} \sum_{i<j} \left(y(x_i) - y(x_j)\right)^2 \tag{10}$$

G. *Evaluation*

Confusion matrix is a measuring tool that can be used to calculate the performance or accuracy of the classification process. It uses a confusion matrix to analyze how well the classifier can recognize records from different classes. Below is a table of confusion matrix used [24].

TABLE V
CONFUSION MATRIX

Actual Label	Positive	TP	FP
	Negative	FN	TN
		Positive	Negative
		Prediction Label	

Based on the existing Table V, it can be calculated the classification performance that was built, among others:

1. Accuracy

Accuracy is the ratio of the number of predictions per document to the number of all predictions classified into these classes [25]. Accuracy can be formulated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\ v \tag{11}$$

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

2. Precision
   Precision is the ratio of the number of relevant documents to the total number of documents found by the classifier[25]. Precision can be formulated as follows:

$$Precision = \frac{TP}{TP+FN} \tag{12}$$

3. Recall
   The recall is the ratio of the number of documents recovered by the classifier to the total number of relevant documents[25]. Recall can be formulated as follows:

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

4. F1-Score
   F1-Score is a combination of the average harmonic of precision and recall, directly proportional to both values [25]. F1-Score can be formulated as follows:

$$F1\ Score = \frac{2(recallxprecision)}{recall+precision} \tag{14}$$

III. RESULT AND DISCUSSION

A. Data processing

This study uses a dataset of 3804 data divided into four labels: 2916 free Flow, 642 stable Stedy Flow, 222 Stable Controlled, and 24 Unstable. In this study, the labeled dataset will be tested, and the dataset will be split into 70% train data and 30% test data by testing each method, namely Support vector machine and Naive Bayes.

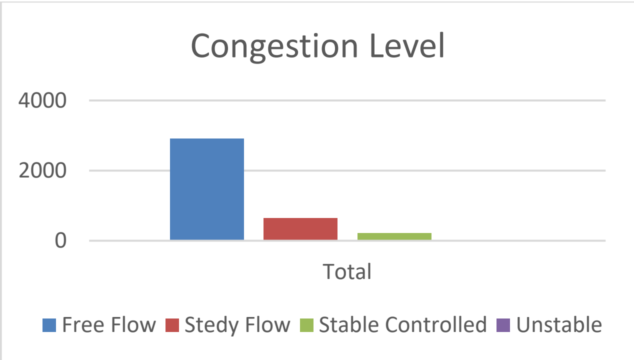

Fig. 2. Data Before Oversampling

Fig. 2. Shows unbalanced data. The classification tends to ignore small classes if the data is not balanced. So that many train data must be included in a small class that is wrongly predicted by the classifier.

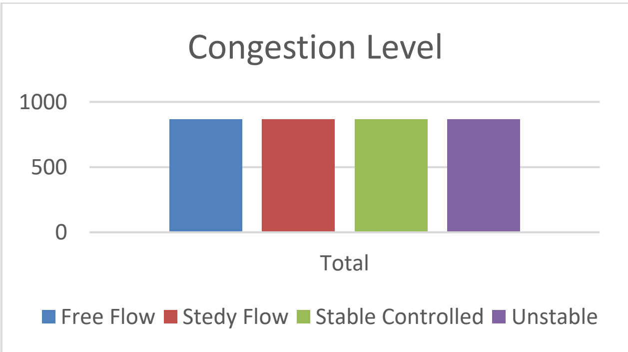

Fig 3. Data After Oversampling

Fig. 3. This shows that the dataset used in this study is balanced using the oversampling technique. Accuracy results obtained from balanced datasets are better than results using unbalanced datasets. Because the data is already

E-ISSN : 2540 - 8984

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Volume 07, Nomor 04, Desember 2022 : 1139 - 1148

balanced, the next step is to perform hyperparameter tuning to get the best parameters, such as Table VI.

TABLE VI
BEST PARAMETER

Model	Parameter	Parameter Value	Best Parameter
Support Vector Machine	C	0.1, 1, 10	1
	Gamma	Scale, Auto	Scale
	Kernel	Linear, RBF, Poly	Linear
Naïve Bayes	Var Smoothing	0, -9	0.0003511191734215131

B. Implementation

In this section, we compare the performance of two methods, Support Vector Machine and Naive Bayes. After the data is classified, the writer conducts trials on each machine learning method. The performance of the method can be seen in Fig. 4, Fig. 5, and Fig. 6.

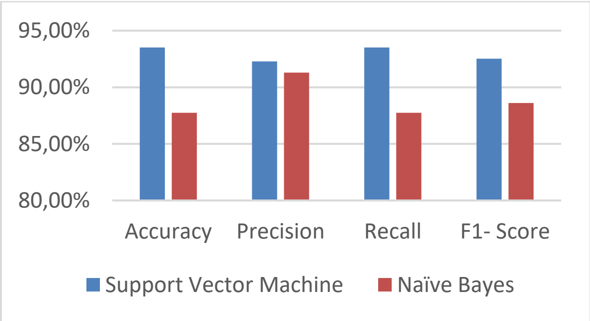

Fig. 4. Result Before Oversampling

Fig. 4 shows the results of the model from Support Vector Machine and Naive Bayes before oversampling. The results obtained for accuracy 93.52%, precision 92.29%, recall 93.52% and f1-score 92.52% with Support Vector Machine and for Naive Bayes get accuracy 87.74%, precision 91.30%, recall 87.74% and f1-score 88.61%.

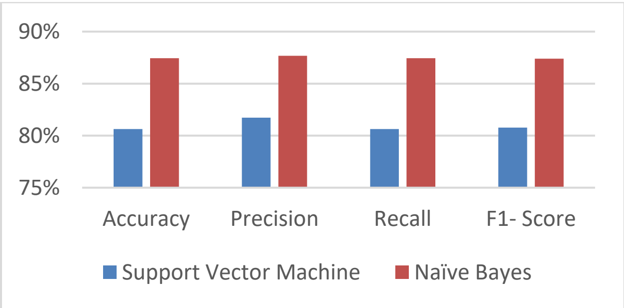

Fig. 5. Result After Oversampling

Fig. 5 shows the model results from Support Vector Machine and Naive Bayes after oversampling. The results obtained with Support Vector Machine accuracy of 88.51%, precision 81.73%, recall 81%, f1-score 80.99% and for Naive Bayes the accuracy is 87.44%, precision 87.68%, recall 87.44& and f1-score 87.40%.

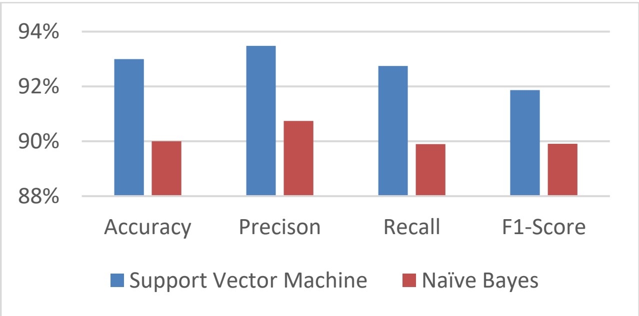

Fig. 6. Result After Hyperparameter Tune

Fig. 6. shows the model results from Support Vector Machine and Naive Bayes with hyperparameter tuning. The results obtained for Accuracy, Precision, Recall, and F1-Score are 93%, 93.48%, 92.75%, 92.68% with SVM and 90%, 90.74%, 89.90%, 89.91% with Naive Bayes.

Fig. IV and Fig. V look to have different results, but after oversampling, the data will be balanced because balanced data is better than using unbalanced data. Unbalanced data can cause poor classification results, so oversampling is needed to improve model performance. Fig. 6. shows that Support Vector Machine has the best accuracy, precision, recall, and f1-score results. From this research, it can be concluded that the Support Vector Machine method performs the classification better.

C. Congestion Classification Map

In making a map, the research is based on a semivariogram dan four-parameter, namely Gaussian, circular, spherical, and exponential. Map making will be done using the best parameter. the following comparison is in table VII.

TABLE VII
RMSE

Model	Variogram Type	RMSE
	Gaussian	0.9256
Support Vector Machine	Circular	0.9046
	Spherical	0.9025
	Exponential	0.8999
	Gaussian	0.9640
Naïve Bayes	Circular	0.9449
	Spherical	0.9481
	Exponential	0.9479

Based on Table VII, the best RMSE can be obtained on the support vector machine model with an exponential type variogram, which is 0.8999, so the congestion map is made using the support vector machine method and ordinary kriging with an exponential type variogram. The following is a map displayed in Fig. 7.

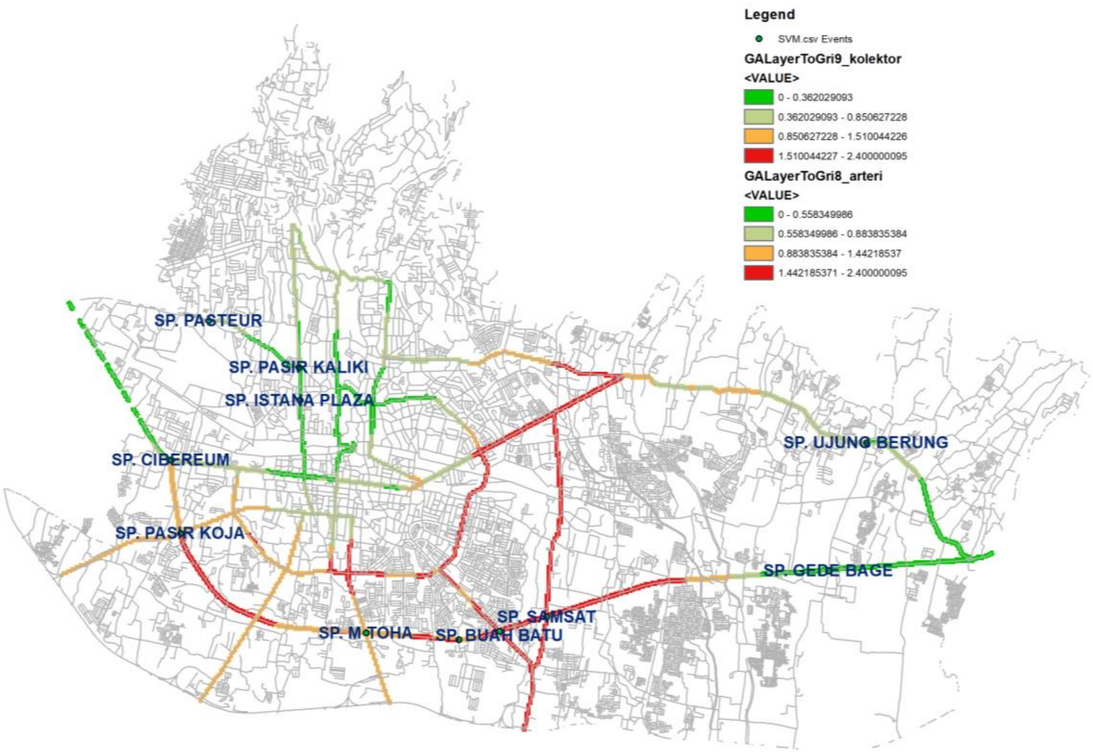

Fig. 7. Congestion Classification Map

Fig. 7 shows that congestion tends to be in the southern area, namely in SP. Samsat and SP. Buah Batu is red, which means it enters the congestion level. Unstable or level 3. And in the western area on the map, namely SP. M. Toha, SP. Pasir Koja, and SP. Cibereum looks orange in color, which means it is in a stable controlled congestion level or level 2. And in the east area of the map, namely SP. Ujung Berung and SP. Gede Bage tends to be dark green which means it enters the level of steady flow congestion or level 1. And in the northern area of the map, namely SP. Pasteur, SP. Pasir Kaliki, and SP. Istana Plaza tends to be light green which means it enters the free flow congestion level or level 0.

he surrounding area. In displaying maps using the ArcMap application, the prediction results of the support vector machine and ordinary kriging methods show good maps and achieve 93% accuracy. Compared with previous studies[6],[7],[8],[9] only show prediction results but do not create congestion maps. Meanwhile, research [10] and [11] only show a map of regional congestion but do not display a road map. In comparison, the results of this study show the results of the classification and the making of a road map.

IV. CONCLUSION

In this study, the authors develop the performance of the support vector machine and the Naive Bayes method to classify traffic jams in the city of Bandung. Three scenarios were tested to see the interpretation of the method in classifying traffic jams in Bandung City with the Bandung City ATCS dataset. Based on the research, the best method for research in two scenarios is the accuracy of the support vector machine. The support vector machine worked well in both scenarios, with 93.52% in the design before oversampling and 93% in the scenario with a hyperparameter setting. However, the accuracy performance of the Naive Bayes method is relatively the same as the accuracy performance of the support vector machine, with 88% in the pre-oversampling scenario and 90% in the hyperparameter setting scenario. Although the support vector engine performs better than Naive Bayes, it takes more time than the Naive Bayes method to train the model. This study found that the support vector machine method is higher than the Naive Bayes method, namely, the support vector machine that uses the best hyperparameter tuning in classifying congestion. The level of accuracy, precision, recall, and f1-score reaches 93%. In making maps using ordinary kriging and the best classification, the support vector machine and support vector machine methods also get the best RMSE with 0.8999 using the exponential semivariogram type. The author uses the best RMSE for map making because using the best RMSE can make the map get a small error result. The result of congestion classification maps is that the southern area of Bandung is denser than other areas of Bandung. This research can be used to predict congestion areas. For further research, other methods can be used to compare the results of other methods and use more data to improve accuracy.

ACKNOWLEDGMENT

The author would like to thank Telkom University and ATCS Bandung city for the support of facilities and infrastructure in providing information and data so that this research can be completed.

REFERENCES

[1] T. Afrin and N. Yodo, “A survey of road traffic congestion measures towards a sustainable and resilient transportation system,” *Sustain.*, vol. 12, no. 11, pp. 1–23, 2020, doi: 10.3390/su12114660.
[2] Q. Zhu, Y. Liu, M. Liu, S. Zhang, G. Chen, and H. Meng, “Intelligent planning and research on urban traffic congestion,” *Futur. Internet*, vol. 13, no. 11, pp. 1–17, 2021, doi: 10.3390/fi13110284.
[3] S. S. Prasetiyowati, Y. Sibaroni, and D. Prabangkara, “The prediction of optimal route of city transportation based on passenger occupancy using genetic algorithm: A case study in the city of Bandung,” *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 3, pp. 1201–1207, 2018, doi: 10.12928/TELKOMNIKA.v16i3.7077.
[4] H. Xiong, A. Vahedian, X. Zhou, Y. Li, and J. Luo, “Predicting traffic congestion propagation patterns: A propagation graph approach,” *IWCTS 2018 - Proc. 11th ACM SIGSPATIAL Int. Work. Comput. Transp. Sci.*, pp. 60–69, 2018, doi: 10.1145/3283207.3283213.
[5] B. Sugara and A. Subekti, “Penerapan Support Vector Machine (Svm) Pada Small Dataset Untuk Deteksi Dini Gangguan Autisme,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 177–182, 2019, doi: 10.33480/pilar.v15i2.649.
[6] Y. R. Tampubolon, S. M. Nasution, R. A. Nugrahaeni, F. T. Elektro, U. Telkom, and D. Indonesia, “BANDUNG MENGGUNAKAN KAMERA CCTV DENGAN ALGORITMA SUPPORT VECTOR MACHINE DESIGN OF APPLICATION FOR DETECTION OF VEHICLE CONGESTION IN BANDUNG USING CCTV CAMERA WITH SUPPORT VECTOR MACHINE.”
[7] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, “Traffic flow prediction based on combination of support vector machine and data denoising schemes,” *Phys. A Stat. Mech. its Appl.*, vol. 534, p. 120642, 2019, doi: 10.1016/j.physa.2019.03.007.
[8] M. T. Zulfikar and Suharjito, “Detection traffic congestion based on twitter data using machine learning,” *Procedia Comput. Sci.*, vol. 157, pp. 118–124, 2019, doi: 10.1016/j.procs.2019.08.148.
[9] J. Yao and Y. Ye, “The effect of image recognition traffic prediction method under deep learning and naive Bayes algorithm on freeway traffic safety,” *Image Vis. Comput.*, vol. 103, p. 103971, 2020, doi: 10.1016/j.imavis.2020.103971.
[10] B. Bae, H. Kim, H. Lim, Y. Liu, L. D. Han, and P. B. Freeze, “Missing data imputation for traffic flow speed using spatio-temporal cokriging,” *Transp. Res. Part C Emerg. Technol.*, vol. 88, no. May 2017, pp. 124–139, 2018, doi: 10.1016/j.trc.2018.01.015.
[11] Z. Yi, X. C. Liu, N. Markovic, and J. Phillips, “Inferencing hourly traffic volume using data-driven machine learning and graph theory,” *Comput. Environ. Urban Syst.*, vol. 85, no. April 2020, p. 101548, 2021, doi: 10.1016/j.compenvurbsys.2020.101548.
[12] E. Alomari, R. Mehmood, and I. Katib, “Road traffic event detection using twitter data, machine learning, and apache spark,” *Proc. - 2019 IEEE*

SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Internet People Smart City Innov. SmartWorld/UIC/ATC/SCALCOM/IOP/SCI 2019, pp. 1888–1895, 2019, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00332.

[13] H. Al Faritzie, "ANALISIS PENGUKURAN DERAJAT KEJENUHAN DAN TINGKAT PELAYANAN RUAS JALAN R . SUKAMTO KOTA PALEMBANG Kota Palembang adalah ibukota provinsi Sumatera Selatan yang merupakan salah satu kota metropolitan di Indonesia . Kota Palembang memiliki luas wilayah 400 ," vol. 6, pp. 131–141, 2021.

[14] 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.

[15] F. Ali, A. Ali, M. Imran, R. A. Naqvi, M. H. Siddiqi, and K. S. Kwak, "Traffic accident detection and condition analysis based on social networking data," *Accid. Anal. Prev.*, vol. 151, no. September 2020, p. 105973, 2021, doi: 10.1016/j.aap.2021.105973.

[16] R. D. Fitriani, H. Yasin, and T. Tarno, "PENANGANAN KLASIFIKASI KELAS DATA TIDAK SEIMBANG DENGAN RANDOM OVERSAMPLING PADA NAIVE BAYES (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal)," *J. Gaussian*, vol. 10, no. 1, pp. 11–20, 2021, doi: 10.14710/j.gauss.v10i1.30243.

[17] J. Golze, S. Zourlidou, and M. Sester, "Traffic Regulator Detection Using GPS Trajectories," *KN - J. Cartogr. Geogr. Inf.*, vol. 70, no. 3, pp. 95–105, 2020, doi: 10.1007/s42489-020-00048-x.

[18] R. Gholami and N. Fakhari, *Support Vector Machine: Principles, Parameters, and Applications*, 1st ed. Elsevier Inc., 2017.

[19] 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.

[20] M. Fajar Rizqi, R. Rendian Septiawan, and I. Kurniawan, "Implementation of Simulated Annealing-Support Vector Machine on QSAR Study of Indenopyrazole Derivative as Anti-Cancer Agent," *2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021*, no. Icidm, pp. 662–668, 2021, doi: 10.1109/ICoICT52021.2021.9527416.

[21] G. Wang and J. Kim, "The prediction of traffic congestion and incident on urban road networks using Naive Bayes classifier," *ATRF 2016 - Australas. Transp. Res. Forum 2016, Proc.*, no. November, pp. 1–14, 2016.

[22] R. A. Wulandari, "Penerapan metode ordinary kriging pada pendugaan kadar karbon monoksida (co) di udara tahun 2020 (studi kasus: pencemaran udara kota surabaya)," *Progr. Stud. Mat. UIN Sunan Ampel*, vol. 2020, 2021.

[23] Q. Jia, W. Li, and D. Che, "A Triangulated Irregular Network Constrained Ordinary Kriging Method for Three-Dimensional Modeling of Faulted Geological Surfaces," *IEEE Access*, vol. 8, pp. 85179–85189, 2020, doi: 10.1109/ACCESS.2020.2993050.

[24] D. J. M. Pasaribu, K. Kusrini, and S. Sudarmawan, "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan Bert Embedding," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 9–20, 2020, doi: 10.35585/inspir.v10i1.2568.

[25] W. C. Widyaningtyas, A. Adiwijaya, and S. Al Faraby, "Klasifikasi Sentiment Analysis Pada Review Film Berbahasa Inggris Dengan Menggunakan Metode Doc2vec Dan Support Vector Machine (svm)," *eProceedings Eng.*, vol. 5, no. 1, pp. 1570–1578, 2018.