ISSN 2278-3091
Volume 9, No.1.5, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse4091.52020.pdf
https://doi.org/10.30534/ijatcse/2020/4091.52020

Traffic Flow Prediction Using Naïve Bayes Based on Measurement of V/c Ratio

Amjad Malik[1], Surya Michrandi Nasution[2], Agus Virgono[3]
[1]Telkom University, Indonesia, amjadmalik@student.telkomuniversity.ac.id
[2]Telkom University, Indonesia, michrandi@telkomuniversity.ac.id
[3]Telkom University, Indonesia, avirgono@gmail.com

ABSTRACT

Along with the rapid selling power of motorized vehicles, both vehicles, motorcycles, and public transportation. The denser the volume of vehicles crossing urban and rural roads, this certainly creates an unpredictable congestion, the process and modeling of traffic flow can be started by studying SUMO (Simulation Urban Mobility). The output value of SUMO is used in calculations in the Python application with the Naïve Bayes algorithm that requires previous history data to be able to predict traffic flow. The input used in making predictions is historical data on the number of vehicles going on Riau Street Bandung. The output generated from the use of the Naïve Bayes method is the level of the jam class that runs on Riau Street Bandung in 2018 used a simulation on the SUMO (Simulation of Urban Mobility) application.Tthe system is based on the existing V/c ratio on the Riau Street of Bandung, because the value of the V/c ratio affects the level of road density. The resulting with Naïve Bayes method 30% Testing data and 70% Training data has an accuracy of 96.61%.

Key words : Traffic flow, SUMO, Python, Naïve Bayes algorithm. V/c ratio.

1. INTRODUCTION

Along with the development of increasingly rapid times, the development of information technology can make it easier for humans to access anything they want. Urbanization is an influence of technological development, where people who live in small areas move to big cities. For example, Bandung, Yogyakarta and Jakarta with the hope of improving the economy [1]. Communities that are urbanizing, make growth in big cities become denser. Population density is very influential on the smooth flow of traffic. The more vehicles that pass through the road, the more congested the traffic flow that passes on the road [2].
Traffic congestion causes a lot of losses for example time loss, if a distance of 64 km can be traveled with 1 hour, then if there is congestion with the same time it may only be able to cover a distance of 10-20 km [3]. Of course, it is very inefficient for people in the era of the development era. Therefore, it is necessary to have a system that can predict the condition of traffic flow in big cities in Indonesia, especially in Bandung.
Of the several problems caused by traffic density in several cities, several traffic flow simulations have been carried out in order to parse density and predict traffic flow in cities, for example by using Simulation Urban Mobility (SUMO), using SUMO we can do simulations modeling of intermodal traffic systems including public transport, pedestrians, and other entities [4]. By simulating the traffic system on the Urban Mobility Simulation (SUMO) application, we can use the traffic history data obtained from the Bandung City Transportation Department, as a reference data to be processed by the system that will be created.
Based on the problems above, a system solution will be made that can predict the flow of traffic on the Riau Street of Bandung using the SUMO application with the Naïve Bayes method. By entering a number of parameters into the system, it is expected that traffic users who will go to the Riau Street of Bandung will be able to know the condition of the traffic flow on the road so that if there is a density, the road users can first know the traffic conditions that they will pass. In order to reduce the level of congestion in the city of Bandung, the Government of the City of Bandung in particular the Bandung City Transportation Agency should work together with various parties to prevent and even reduce the level of congestion in the city of Bandung.

2. LITERATURE STUDY

2.1 Level of Service (LOS)

Level of Service (LOS) is a benchmark to assess the performance of a road section. Determination of an LOS based on the value of volume / capacity ratio (V/c ratio). If the V/c ratio is getting closer to 1 then the performance on a road segment will be lower. Each road section has a different V/c ratio[5].

Table 1: V/c Ratio

LOS	Range V/c Ratio
A	0 - 0.24
B	0.25 - 0.45
C	0.46 - 0.76
D	0.77 - 0.87
E	0.88 >

V/c ratio is the number of vehicles on one road segment at a time compared to the capacity of the highway, the value of V/c ratio is determined in decimal eg 0.8 or 1.2 if the value of the V/c ratio is less than 1, then the road smooth traffic, if equal to 1 means the traffic on the road is in accordance with its capacity, and if more than 1 means the traffic is congested or congested. The value V/c ratio c also determines the Service Level (LOS) or road service level which is denoted by the letters A to F, while A (Current Vehicles) and F (Very Traffic)the value V/c ratio c also determines the Service Level (LoS) or road service level which is denoted by the letters A to F, while A (Current Vehicles) and F (Very Traffic Jam).

2.2 Simulation of Urban Mobility (SUMO)

SUMO (Simulation of Urban Mobility) is a series of open, microscopic, integrated traffic simulations designed for the safety of large-scale road networks[6]. SUMO can model intermodal traffic systems including public transport and pedestrians. Many tools can be used in SUMO applications, such as route search, visualization, traffic flow modeling and gas vehicle emission calculations[7].
SUMO can be improved with special models to control simulations from remote control. SUMO can work by creating its own traffic manually or can import maps from Open Street Maps (OSM).

2.3 Traffic Flow

Traffic flow is running back and forth, going back and forth and about traveling on the road and so on and related between one place with another place. Whereas mentioned in Law No.22 of 2009, traffic is defined as the movement of vehicles and people in the road traffic space. Traffic space itself is infrastructure in the form of roads and supporting facilities intended for moving vehicles, people or goods. In traffic, there are 3 (three) component systems which include humans, vehicles and roads that interact with each other in the movement of vehicles[8].

2.4 Open Street Map (OSM)

OpenStreetMap (OSM) is a tool for delivering a free and open world map, built by some volunteers who survey the geographic area by using GPS, digitize aerial imagery, and collect and free public sources available.
OSM is using the Open Data Commons Open Database License 1.0. OSM contributors not only can own and modify data, but also share map data to the public. There are lots digital maps option that available over the internet, but most of them have strict restrictions in it license, which makes difficult to others who worked for governments, or for researchers and academics, innovators, and others to use it free and open to use the data that available on the maps. On the other hand, OSM gives both the base map and the data which can be used for further development and redistribution.
In many parts of the world, especially in remote and backward areas, there is no incentive for mapping efforts to develop data in such places. In this case, OSM is an alternative for economic development, emergency response, urban planning, and many other objectives[9].

2.5 Naïve Bayes Algorithm

Naïve Bayes algorithm is one of the classification methods that use probability and statistical methods to predict a situation [10-12]. This method proposed by British scientist Thomas Bayes. The Naïve Bayes algorithm predicts future based on its past experience, and this methods also known as the Bayes Theorem. The main characteristic of Naïve Bayes Classifier is giving a very strong assumption (Naïve) of independence from each condition [13]. Figure 1 shows the implementation of the Bayes Classifier for defining traffic prediction.
Bayes' theorem has the following general form[14]:

$$P(H|X) = \frac{P(X|H)}{P(X)}.P(H)$$

X : Training Data (unknown classes)
H : The hypothesis data for a specific class
P(H|X) : The probability of hypothesis (H) is based on condition (X)
P(H) : probability Hypothesis (H)
P(X|H) : Probability of Data (X) is based on its hypothesis (H)
P(X) : Probability of Data (X)

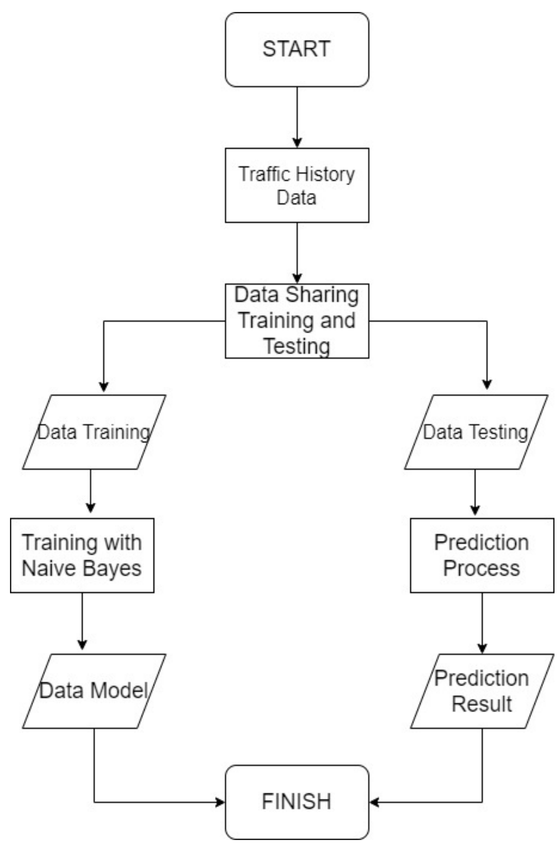

Figure 1: Naïve Bayes Algorithm Flowchart

On the prediction process is done with the following steps. Data input in the form of traffic history data and several parameters is then carried out Distribution of training data and testing data, After the distribution of testing data and training data then the process continues into training using the Naïve Bayes method and the distribution of testing data will continue to the prediction process that is added to the result model data training with the Naïve Bayes method. Then the output of this system is in the form of prediction results which are divided into several classes namely A, B, C, D based on the level of traffic density on Riau Street, Bandung City.
Level of Services (LOS) is a benchmark to assess the performance of a road section. Determination of a LOS based on the value of volume/capacity ratio (V / C ratio). If the V / C ratio is getting closer to 1 then the performance on a road segment will be lower. Each road section has a different V / C ratio [15]. The type of LOS used in this paper is the Level of Service depending on the flow. This level of service assesses a road segment based on vehicle speed on a road or road facility [16].

3. SOME COMMON MISTAKES

3.1 System Overview

The system that will be built in this project is a simulation using the SUMO application of traffic flow on Riau Street, Bandung City, which is expected to be able to help related agencies, especially the Bandung City Transportation Office in the management and regulation of traffic flow on the road. In this study the data will be processed using the naïve Bayes method. The case discussed in this study is the prediction of traffic flow on Riau Street, Bandung.

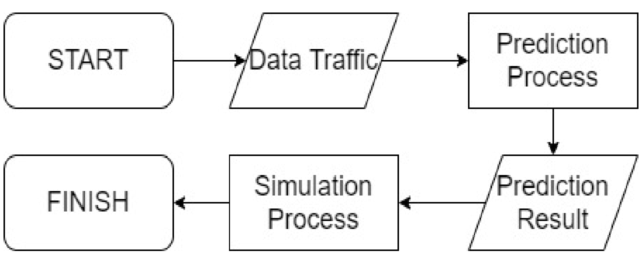

Figure 2: General Description System

In Figure 2 the input data that will be provided for traffic flow prediction is the traffic history data in the form of a CSV file obtained from the Department of Transportation of the City of Bandung. The features used to predict traffic flow include day, hour, minute, number of vehicles, vehicle type, and road V/c ratio. Then the input data will be tested and trained on a python application so that it will output the data that has been processed using the naïve Bayes method and also the input data is implemented and simulated on the SUMO application. The purpose of building this simulation system is to be able to predict the traffic flow on Riau Street, Bandung City so that road users are expected to know in real time the conditions of the traffic flow on the road.

3.2 Simulation Process using SUMO Application

In the process of traffic simulation in the SUMO application is done with several steps, namely the first stage to carry out a simulation is to collect the required data, such as the map of Riau Street, Bandung, as well as the traffic data on the road. To get map data on Riau Street, Bandung City, you can go through the OpenStreetmap website. Then Maps that have been obtained in OpenStreetmap must be changed to the format .net.xml. this is intended so that the SUMO application can read the map. After the map file has been converted to the .net.xml format, route creation can be done. In making this route file can be done in several ways, one of which is using tools from the SUMO application, randomtrips.py. the output of this route determination process is a file with the format ".rou.xml". Figure 3 shows the process of the simulation using SUMO.

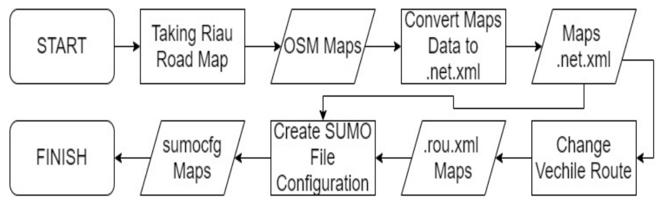

Figure 3: Simulation Process with SUMO Application

The first process carried out is the creation of a system for the prediction process. engine created using the naïve Bayes method. In making engines with the naïve Bayes method, we must pay attention to the distribution of data to be used. Because if you do not pay attention to the spread of data, the engine accuracy results are not good. If the data distribution is known, then it can determine the proper use of the kernel.
After making the engine is complete the process to split the dataset into training data and test data can be done. In naïve Bayes the more training data, the better the prediction results. At the training data processing stage, determining the kernel is very important. The kernel used must be able to create hyperplanes that share data well. So, the data that has been divided into classes does not have a lot of noise. When this process is complete, the data will be stored for testing the test data

4. RESULT AND DISCUSSION

In testing using the Naive Bayes algorithm the data is divided into two, training data and test data. From data sharing, the performance of the Naïve Bayes algorithm will be measured.

Table 2: Naïve Bayes Performance

No	Data Trainin	Data Testing	Performance
1	50%	50%	96,87%
2	60%	40%	96,87%
3	70%	30%	96,61%
4	30%	20%	97,46%
5	90%	10%	96,48%

With the data used and processed using the Naive Bayes algorithm, the following is the distribution of two-way traffic data on the Riau Street in Bandung.

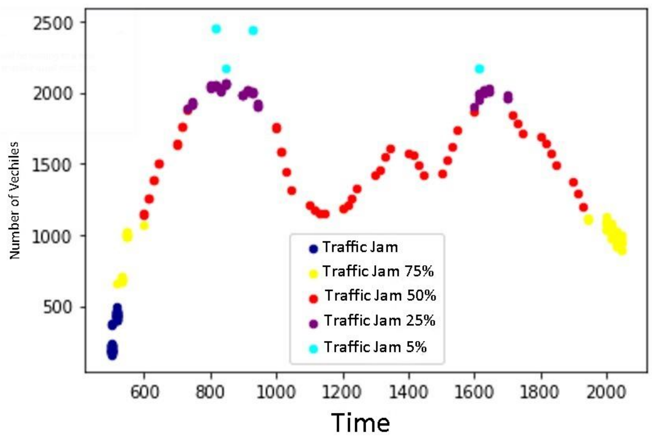

Figure 4: Dissemination of traffic data

Scenario testing simulation of traffic flow on Riau Street, Bandung City. Here the prediction test will be carried out on weekdays and the time is at 10.00 WIB with the number of vehicles 540.

Table 3: Result Traffic Flow Prediction

No	Day	Time	Number of Vechile	Class
1	Monday	10.00 AM	540	C
2	Tuesday	10.00 AM	540	C
3	Wednesday	10.00 AM	540	B
4	Thursday	10.00 AM	540	B
5	Friday	10.00 AM	540	B

4.1 V/c Ratio Simulation Result

a. Class A Simulation

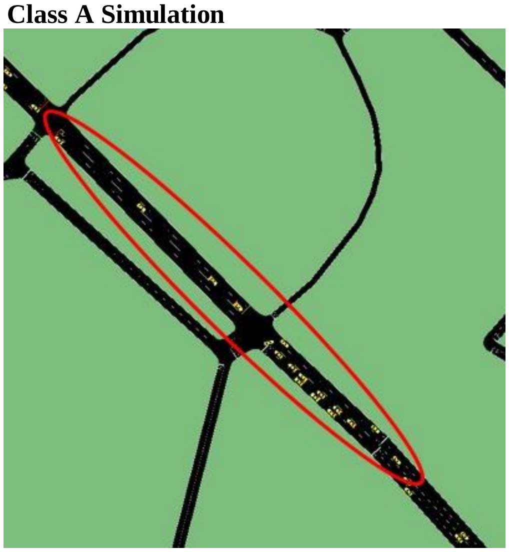

Figure 5: Class A Simulation Result

b. Class B Simulation

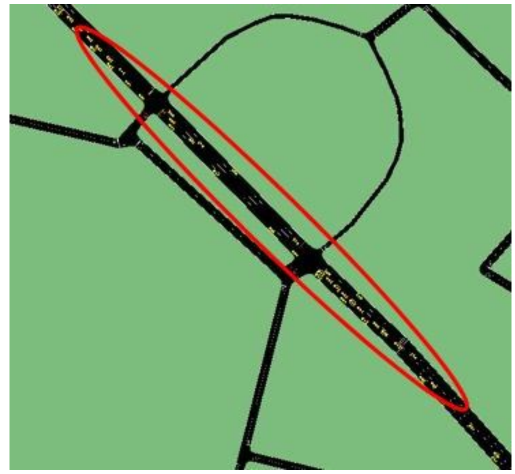

Figure 6: Class B Simulation Result

c. Class C Simulation

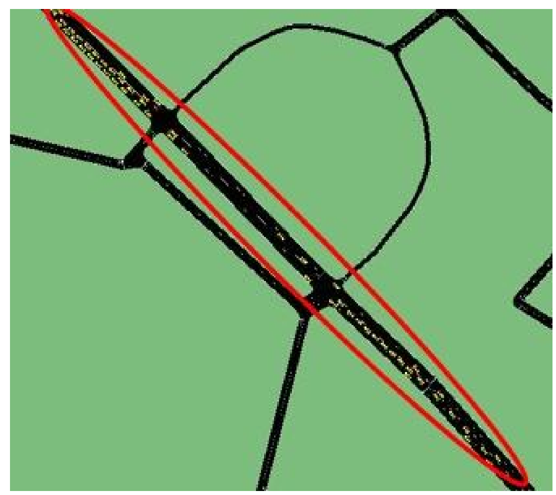

Figure 7: Class C Simulation Result

d. Class D Simulation

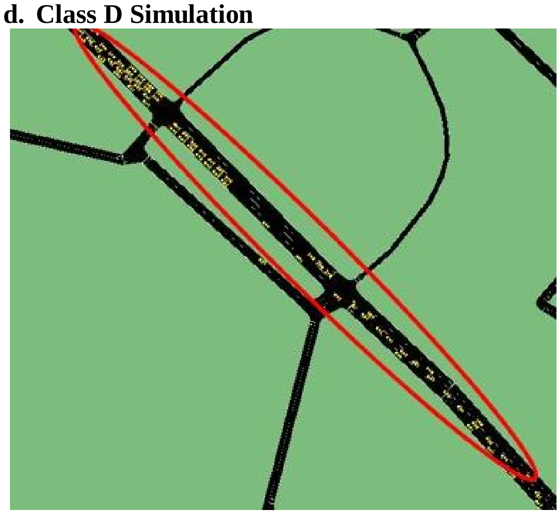

Figure 8: Class D Simulation Result

e. Class E Simulation

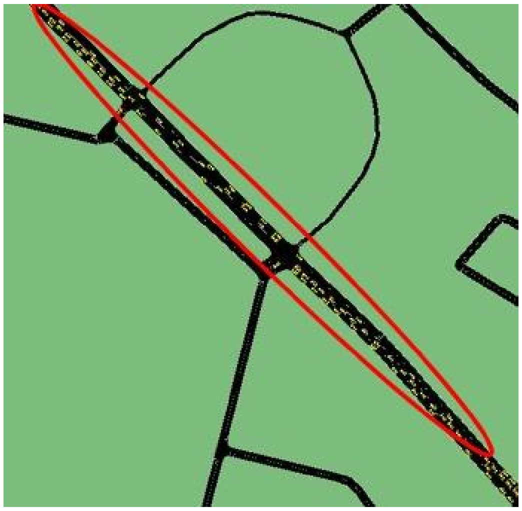

Figure 9: Class E Simulation Result

4.2 Average Speed Calculation

From the average speed level of service graph above, X shows speed, and Y shows class it can be seen the average speed of each class issued. Shows that the level of speed of class A vehicles is at number 18 while class E is at number 10. Because if the level of traffic density decreases, it is also followed by the speed of vehicles passing through the road. Seen in the graph from class A to class E the line continues to decline.

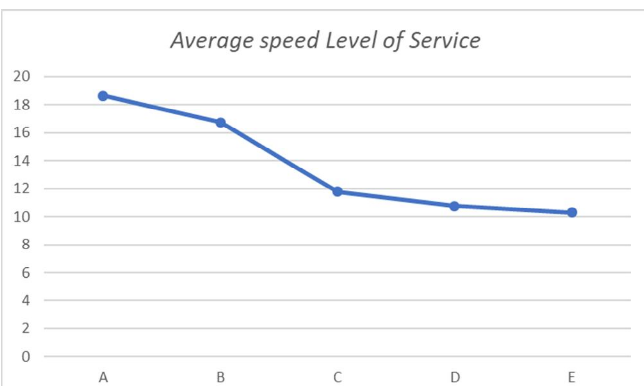

Figure 10: Average Speed Level of Service

The decision boundaries give a result that the classification process can make a good hyperplane to separate every class. This is the reason why the accuracy nearly perfect. But it gives a glimpse that the data has an overfitting condition.

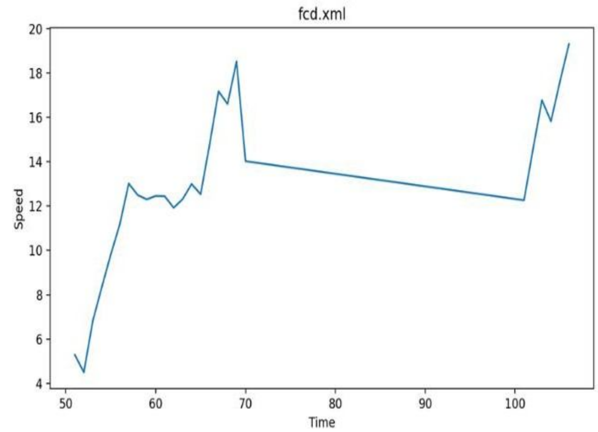

Figure 11: Speed VS Time of LOS A

Can be seen if the LOS A time is high because the speed traveled on the road is not too bad.

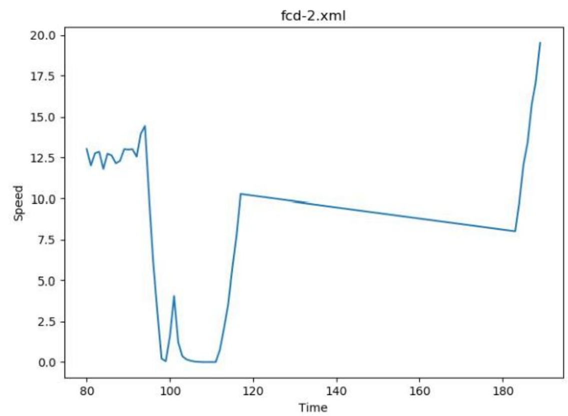

Figure 12: Speed VS Time of LOS B

LOS B graph looks good average travel time, because traffic density does not increase.

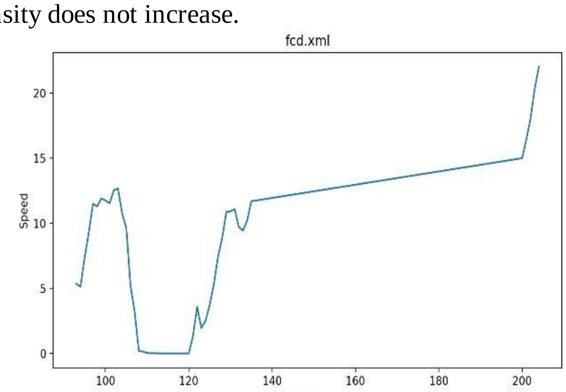

Figure 13: Speed VS Time of LOS C

The average speed is slightly up, because of the ever-increasing traffic density, but it hasn't experienced traffic jams yet.

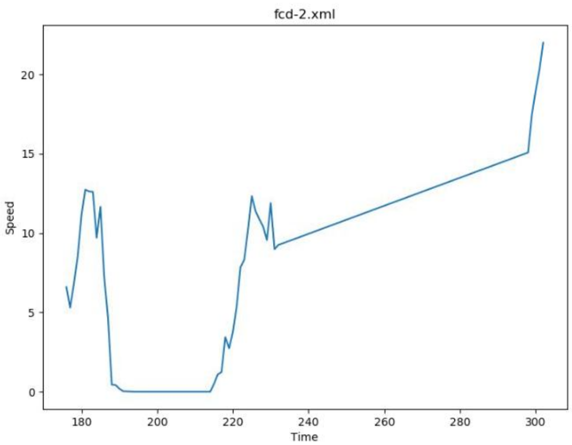

Figure 14: Speed VS Time of LOS D

The travel time for LOS D is longer, because the number of vehicles that pass has caused traffic jams.

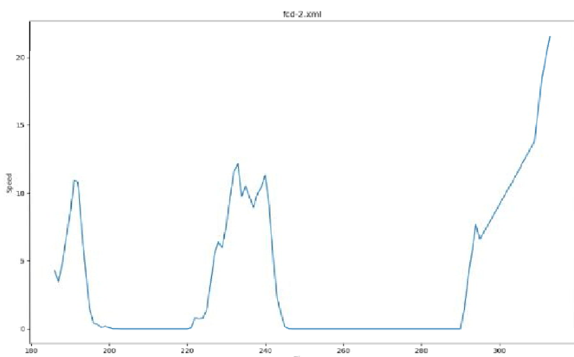

Figure 15: Speed VS time of LOS E

LOS E is the longest travel time, due to the increasing number of vehicles and causing congestion on roads.

5. CONCLUSION

From the results of the system, it can be concluded on the distribution of traffic history data on Riau Street, Bandung City, it can be seen the busy time and the increasing traffic flow density on sections A to B and vice versa, then the predicted class value depends on the volume of the vehicle and also V/c ratio on Riau Street, Bandung City. and the last based on the results of the prediction system performance can be seen that the score issued by the naïve Bayes method is 96.6%.

REFERENCES

1. Muhammad, H. (2017). Penyebab Kemacetan Lalu Lintas di Kota Besar. Jakarta: Replubika.
2. Angriani, P. (2014). Dampak-Dampak Urbanisasi. Jakarta: Kompasiana. Available online at: https://www.kompasiana.com/puspitaanggriani/54f927e1a3331142038b4701/dampakdampak-urbanisasi.
3. BN, A. (2018). Dampak Kemacetan Bagi Kota Besar. TribunNews. Available online at: https://www.tribunnews.com/tribunners/2016/07/21/kemacetan-di-kota-kota-besar-indonesia.
4. Kumiko Maeda, A. U. (2009). Urban pedestrian mobility for mobile wireless. Tokyo: ScienceDirect.
5. Direktorat Jenderal Bina Marga, Manual Kapasitas Jalan Indonesia, Jakarta: Bina Karya, 1997.
6. Daniel Krajzewicz, Georg Hertkorn and Peter Wagner, 2002. "SUMO (Simulation of Urban Mobility) An open-source traffic simulation", Institute of Transportation Systems German Aerospace Centre.
7. Michael Behrisch, Laura Bieker, Jakob Erdmann, Daniel Krajzewicz, 2011. "SUMO - Simulation of Urban Mobility", Institute of Transportation Systems German Aerospace Centre.
8. Argya Sukma Jiwangga. "Analisis Faktor Pengaruh Lalu Lintas". Fakultas Teknik UMP, 2017.
9. Kramer, J. (2011). “What Is OpenStreet Maps ?”. London UK: derickrethams.nl.
10. Kareem, S.W., Mehmet Cudi Okur, “Bayesian Network Structure Learning Based On Pigeon Inspired Optimization”, International Journal of Advanced Trends in Computer Science and Engineering, 8(1.2), 2019.
11. Jayadi, R., Hafizh M.F., Muhammad T.J.D, Muhammad F.S., Alfitra M.P., “Employee Performance Prediction using Naïve Bayes”, International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 2019.
12. Christ M.J., Rahmanto Nikolaus Permana Tri, Wiranto Chandra, Tuga Mauritsius, “Lending Club Default Prediction using Naïve Bayes and Decision Tree”, International Journal of Advanced Trends in Computer Science and Engineering, 8(5), 2019.
13. Alfa Saleh, "Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa", Konferensi Nasional Pengembangan Teknologi Informasi dan Komunikasi. STMIK Potensi Utama, Medan. 2012.
14. Seongwook Youn, Dennis McLeod, "A Comparative Study for Email Classification".
15. Direktorat Jenderal Bina Marga, Manual Kapasitas Jalan Indonesia, Jakarta: Bina Karya, 1997.
16. O. Z. Tamin, Perencanaan dan Pemodelan Transportasi, Bandung: Penerbit ITB, 2000.