Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Statistical Model in Predicting Traffic Congestion Among Selected Routes in Metro Manila

Ruth E. Montes[1*], Laurence P. Usona, MS[2]

[1*,2]*College of Science, Polytechnic University of the Philippines.*

Corresponding Email: [1]rmontes.works@gmail.com*

Received: 27 February 2023 **Accepted:** 22 April 2023 **Published:** 01 June 2023

Abstract: Traffic congestion is a serious issue that contributes significantly to economic loss, increase in greenhouse gas emissions, and fuel wastage. Hence, an accurate congestion prediction model can help address these problems. This paper analyzes the status of the road transport infrastructure, public transportation system, volume of vehicles, road crash data, and government policies, rules, and regulations, as well as the quality of implementation. Moreover, traffic congestion prediction models were developed, using logistic regression, random forest, and neural networks. Seventeen months of daily traffic data were used in developing the models. Results showed that the Random Forest models have recorded the highest accuracy (77%), recall (77%) and F1-score (77%). On the other hand, the Neural Network model has better performance in predicting Free Flow traffic congestion at 81% F1-score, while the Random Forest model showed better results in predicting Moderate, Heavy, and Standstill Traffic.

Keywords: Traffic Congestion Prediction Models, Logistic Regression, Random Forest, Neural Networks, Traffic Data.

1. INTRODUCTION

Road transportation makes up the majority of passenger kilometer traveled within the country and a vital part of the daily lives of every Filipinos. Based on the 2010 study of the World Bank, road sector accounts for 98% of passenger traffic from all sector of transportation and 58% of all cargo traffic (Dato et. al, 2010). Over the years, the Land Transportation Office recorded a 96% increase in Motor Vehicle Registration from 2010 to 2021(LTO Annual Report, 2010 and 2021). However, road capacities have not increased significantly to accommodate the rapid growth in road traffic (Santos, 2020). This leads to road traffic congestions, especially in highly urbanized areas.
Traffic congestion is characterized by progressive reduction in travel speeds, which results into increase in travel times, fuel consumption, and greenhouse gas emissions. The TomTom

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

International B.V., a location technology company, developed an index which provides comprehensive insights about levels of traffic congestion in more than 400 cities all over the world. According to the TomTom Index, Metro Manila ranked second in 2019 with drivers expecting to spend an average of 70% extra travel time being stuck in traffic. This means that travelers spent an extra time of 29 minutes in traffic for every 30-minute trip in the course of morning rush hour and an extra 38 minutes for every 30-minute trip in the evening (TomTom Traffic Index, 2019).

Moreover, according to the projections of the Japan International Cooperation Agency (JICA), 3.5 billion pesos are lost daily because of congested roads in Metro Manila and the economic loss could reach 5.4 billion pesos a day by 2035 (Ochave, 2022). According to the public transit sub-index of the Urban Mobility Readiness Index (UMRI) 2022, the public transportation system in Metro Manila is ranked as the fifth worst in the world. This is based on how Metro Manila compares to other cities across the world. The statistics measured by public transportation which are density, efficiency, and usage rate, as well as the degree to which they can adapt to address competition from emerging mobility services, were included in the calculation of the index.

Predicting traffic congestion is a very challenging task since it has a lot of time-varying characteristics together with many uncertain factors like weather, time, and day (Deshpande and Bajaj, 2017). Hence, a predictive model incorporating all the given factors should be developed. For this study, a long-term approach would be used to be able to predict traffic conditions for weeks and months by developing and comparing the results of three statistical models: multinomial logistic regression, decision trees, and neural networks.

This study aims to develop a traffic analysis and prediction model as a tool in solving road congestions and other traffic related problems, and adopting new transport mechanisms, plans, and programs to improve traffic management and control system in Metro Manila. Specifically, it sought to answer the following questions:

1. What is the traffic situation on the selected routes in terms of the following.

1.1. Route Distance;
1.2. Travel Time;
1.3. Average Speed;
1.4. Congestion Level;
1.5. Day of the week;
1.6. Time of the day and;
1.7. Weather Condition?

2. What is the accuracy, precision, and recall of the following traffic congestion prediction models:

2.1. Multinomial Logistic Regression
2.2. Multiclass Decision Forest
2.3. Multiclass Neural Networks

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

3. Which of the prediction models could best predict traffic congestion in terms of accuracy, precision, and recall?

2. METHODS

The data used in this study was extracted from the Department of Transportation Metro Manila Traffic Watch which displays traffic data along the 31 rationalized bus routes in Metro Manila as published by the Land Transportation Franchising Regulatory Board (LTFRB) and weather data recorded from the Science Garden, Weather Station in Quezon City.

The data consists of the following information:

1. Route Information: Contains details about the route, the direction of the route and the average length of the specific route
2. Trip Information: Contains the details about the speed of the vehicle during the time of the trip. The date of the trip was also recorded in the database; hence the day and month of the trip can be extracted. Congestion levels were also recorded during the time of the trip.
3. Weather data: temperature, wind speed, humidity, and visibility.

In this study, the input data were the route, route length, day, travel time, speed, type of road, and weather data. Three statistical modeling techniques were applied in the dataset to predict traffic congestion for each route at a specific given time. Models were evaluated according to their classification accuracy, precision or sensitivity, recall, and F1-score.
Congestion level was based on the traffic jam information that includes real time data about traffic slowdowns on a specific road segment. Traffic jam information was generated based on the GPS location points sent from user phones and calculations of actual speed against average speed on a specific time and free flow speed which was the maximum speed measured on the road segment.

Descriptive analysis of the selected data was also done using different visualization and spatial temporal maps. In building the models, the data was divided into training data and test data. Categorical variables were coded to numerical values. The training data was run using the algorithm process in Python and returned the predicted output. A comparison was made between the projected result and the actual data. The comparison result of the algorithms will reveal how accurate the prediction is at the final stage. The methods used in building the model are the multinomial logistic regression, random forest, and multiclass neural networks.

3. RESULTS

Traffic Situation on the Selected Routes

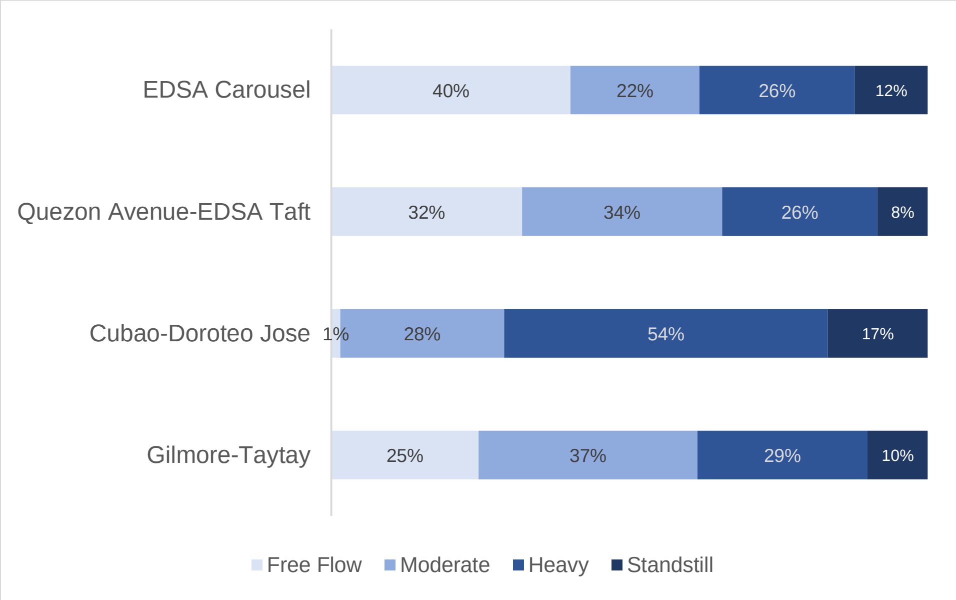
Route Distance

Among the selected routes, Cubao – Doroteo Jose and Edsa Carousel have the longest distance covering 19 km followed by Quezon Avenue – EDSA Taft with 17 km, and the shortest route, Gilmore – Taytay at 14 km. Cubao – Doroteo Jose has the most share of heavy and standstill traffic, while Quezon Avenue – EDSA Taft has the least share of heavy and standstill traffic.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Hence, standstill and heavy traffic were more frequent in Cubao – Doroteo Jose followed by Gilmore – Taytay.

Figure 1 Distribution of Traffic Congestion Types by Route



Free-flow traffic was very seldom at the Cubao-Doroteo Jose route, with only 1% share in comparison with the occurrence of all traffic congestion levels in the specific route.

Travel time

The total amount of time traveled, as determined by this study, reveals that Sundays always comprise the least amount of time traveled, whereas Fridays typically involve the longest time. Monday through Thursday sees a progressive decline in time travel, with Fridays seeing a sharp increase. The highest average travel time recorded was at on a Friday at EDSA Carousel North Bound Route.

EDSA Carousel

The longest travel time in EDSA Carousel takes more than 80 minutes in the North bound route direction and takes more than 60 minutes in the south bound route every Friday and both route directions only take around 50 minutes every Sunday. It can also be seen that travel time for the North Bound route was usually longer compared to the South Bound route, except on a Sunday.

Quezon Avenue – EDSA Taft

The East and West bound routes for Quezon Avenue to EDSA Taft has a minimal difference in travel time. The longest travel time occurs every Friday with almost 20 minutes in the West Bound direction and around 18 minutes for the East Bound direction. Average travel time for the West Bound route was a bit longer than the East Bound for all days of the week.

Cubao – Doroteo Jose

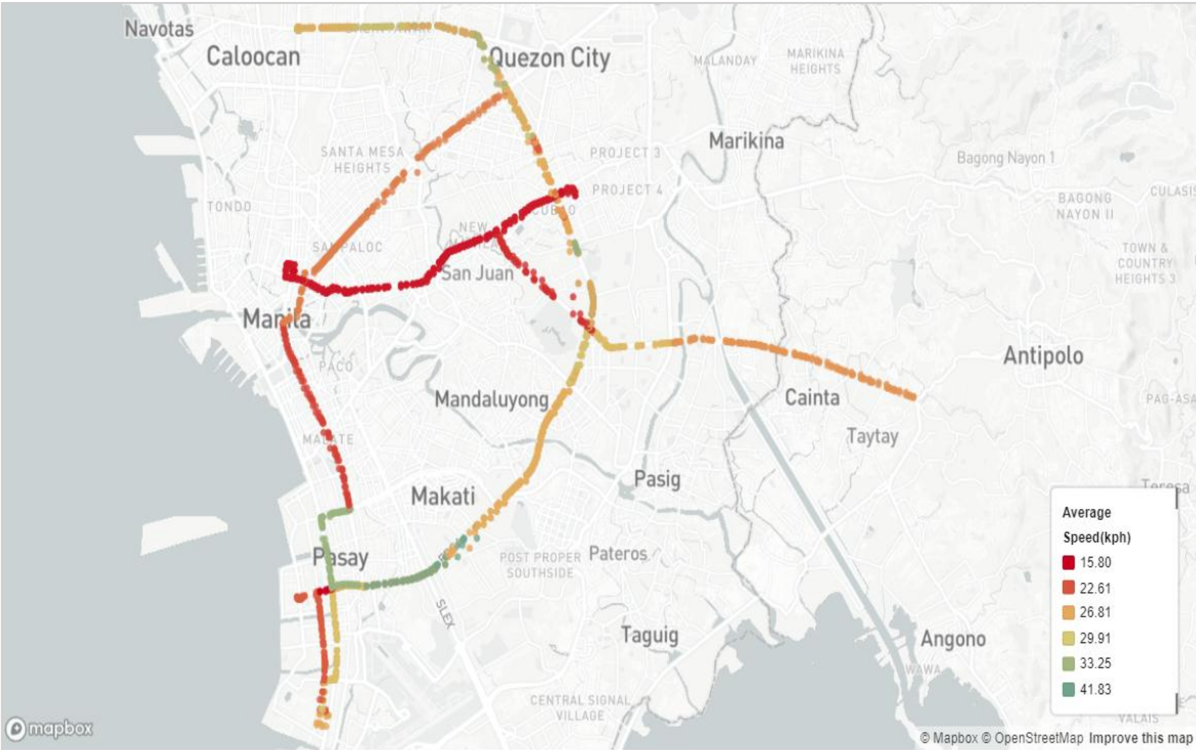
Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

The route between Cubao and Doroteo Jose takes the least amount of time to travel on Sundays, with a difference of almost 5 minutes on Mondays, Tuesdays, Wednesdays, Thursdays, and Saturdays. And the longest time takes more than 25 minutes for the East bound direction and approximately 21 minutes for the West bound direction during Friday.

Gilmore – Taytay
The Gilmore – Taytay West bound route takes a roughly 2 minutes longer than the East bound route with the longest time travel reaching more than 14 minutes during Friday in both East and West Routes. West bound route also recorded a consistent longer average travel time compared to the east bound route for all days of the week.

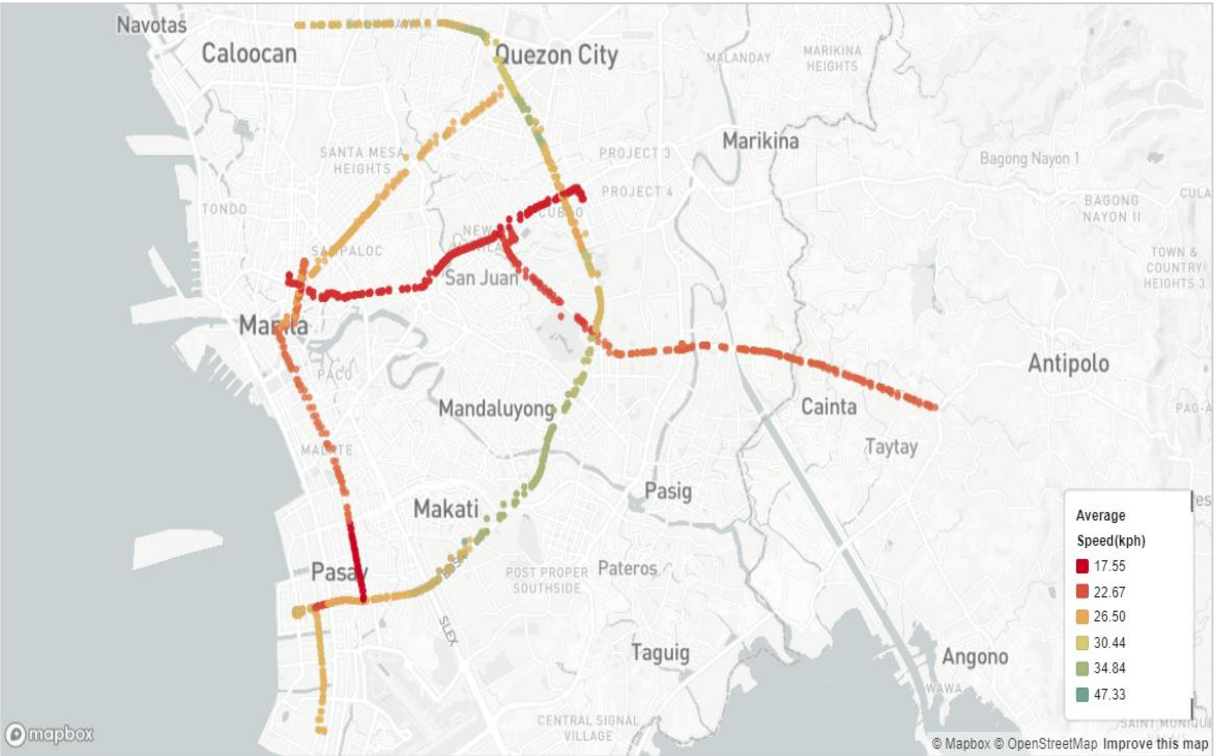
Daily Average Speed
EDSA Carousel regardless of direction exhibits the lowest average speed when compared to all other routes. Looking at the average daily speed, northbound and eastbound routes have lower speeds compared to the southbound and westbound routes. The lowest average speed for the north bound and east bound routes was at 15.80 kph, while the highest was recorded at 41.83 kph at EDSA-Carousel route. The same trend was observed for the opposite routes, although the recorded lowest average speed was a bit higher at 17.55 kph and the highest speed at 47.33 kph.

Figure 2. Mapping of Daily Average Speed for North Bound and East Bound Routes



Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Figure 3 Mapping of Daily Average Speed for South Bound and West Bound Routes



EDSA Carousel

The highest average speed in Edsa Carousel reached above 40 kph at around 12:00 AM that gradually decreased to below 30 kph in 6:00 PM for both North and South bound routes. South bound routes recorded higher average speed when compared to the North Bound route. The largest difference in the average speed between the routes was recorded at 5:00 PM.

Quezon Avenue – EDSA Taft

For the Quezon Avenue – EDSA Taft route, the East bound average speed was usually higher than the West bound route, except for 6:00 PM, where the speed for both directions were close to each other. The highest recorded average speed was around 46 kph at the South Bound route, which happened at 2:00 AM, while for the opposite route the highest average speed was also recorded in the same time at around 43 kph.

Cubao – Doroteo Jose

Cubao – Doroteo Jose recorded the highest average speed at around 25 kph at 4:00AM at the west bound route. For the east bound route, the highest average speed was recorded on 11:00 PM at around 22 kph. Lowest average speed for the west bound route was recorded at around 17 kph at 8:00 AM and around 13 kph for the opposite route which was recorded at 5:00 PM. Average speed in the west bound route was usually higher than the east bound route except between 6:00 AM to 9:00 AM.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Gilmore – Taytay

The average speed of the east bound direction of the Gilmore – Taytay route were generally higher compared to the opposite direction, except between 4:00 PM to 9:00 PM. The highest average speed for the east bound route was recorded at 3:00 AM with an average of around 34 kph, while for the west bound route the highest average speed was recorded at 12:00 AM with an average of around 32 kph. For the lowest recorded speed, the west bound direction recorded an average of around 17 kph at 3:00 PM and for the opposite direction at 15 kph at 5:00 PM.

Congestion Level

EDSA Carousel

In EDSA Carousel route, the north bound direction is seen to have more frequent standstill traffic compared to the south bound direction. It can also be noted that heavy and standstill traffic occurs more often in the afternoon, at around 2:00 PM to 6:00 PM. For the south bound route, it can be observed that standstill traffic was recorded at 6:00 AM to 12:00 NN in the morning and 3:00 PM to 7:00 PM in the afternoon.

Quezon Avenue – EDSA Taft

In Quezon Avenue – EDSA Taft route, the east bound direction started to experience heavy traffic at 6:00AM. Most heavy traffics were reported between 12:00NN to 5:00 PM, while for standstill traffic were recorded between 2:00 PM to 6:00 PM. For the west bound direction, it is noted that heavy traffic was reported all throughout the day. Free-flow traffic was least experienced in this direction.

Cubao – Doroteo Jose

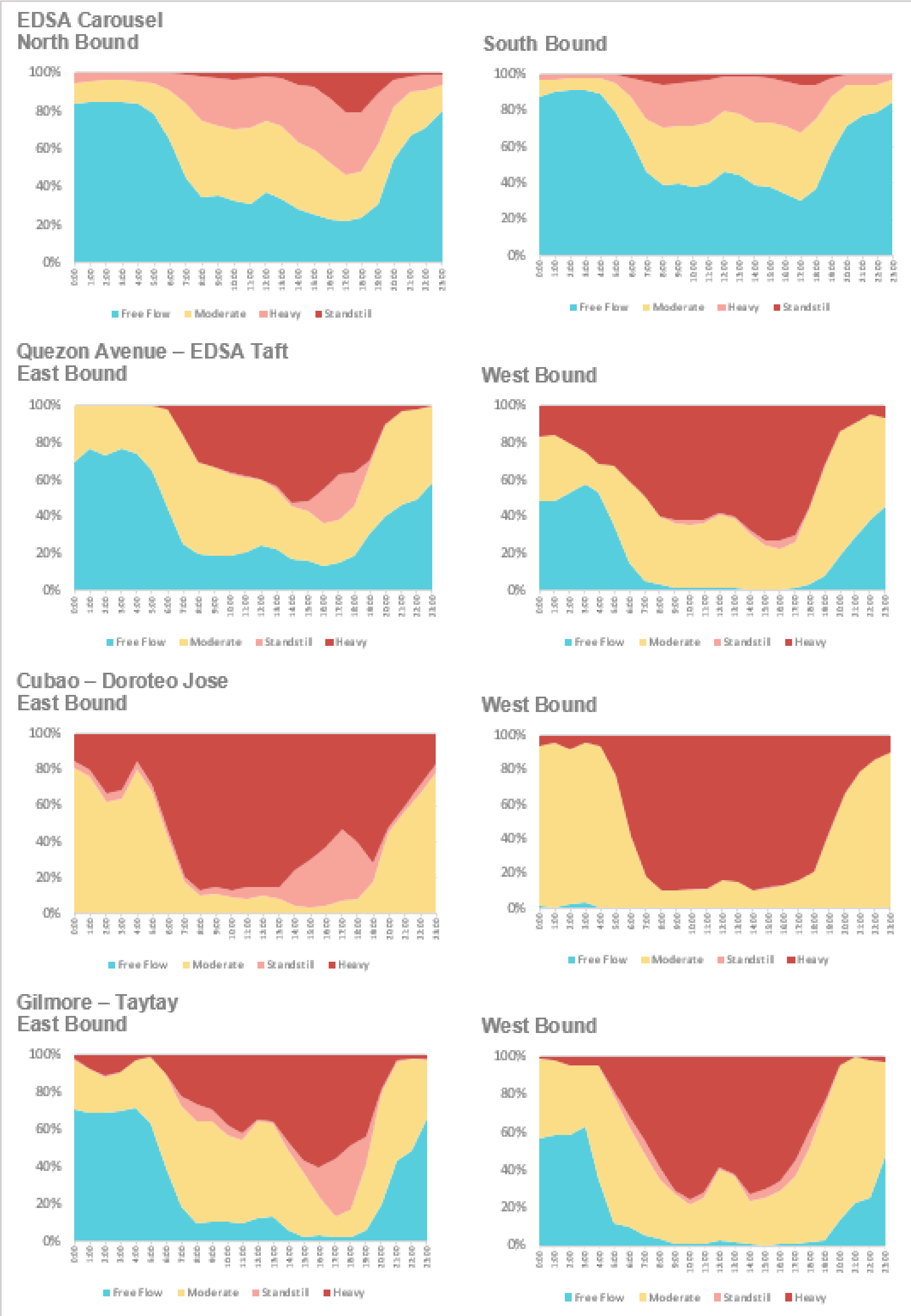
Data showed that free-flow traffic was very seldom at the Cubao – Doroteo Jose route. It can be noted that heavy traffic was more frequent for both directions. For the east bound route, it is noted that standstill traffic peaks at 4:00 PM to 5:00 PM. On the west bound route, it is noted that some free-flow traffic was recorded from 12:00 MN to 3:00 AM.

Gilmore – Taytay

The congestion levels of both west and east bounds in the Gilmore – Taytay route were mostly a mix of moderate and heavy traffic and free flow mostly during the early morning and late evening. The west bound route has longer duration of heavy traffic levels but for east bound, it has more standstill situation.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Figure 4 Congestion Levels by Route and Time

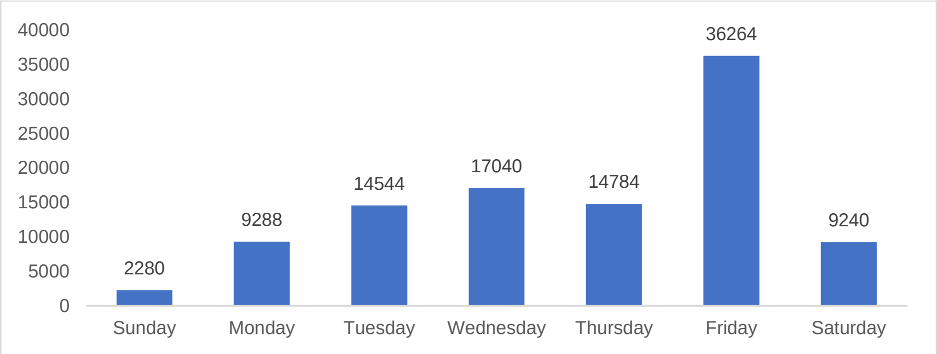


Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Day of the week

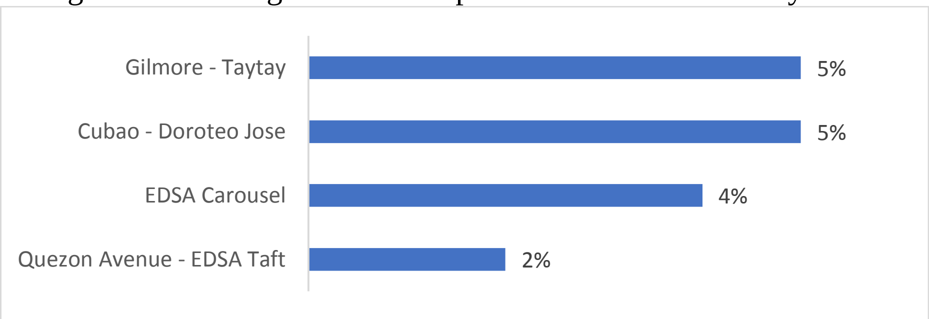
It can be identified from the data gathered in this study that standstill traffic is dominant during Fridays in these routes and is followed by Wednesday and traffic is free flowing every weekend.

Figure 5 Frequency of Standstill Traffic by Day



Looking at the distribution per route, Cubao – Doroteo Jose recorded the most share of standstill traffic, while Quezon Avenue – EDSA Taft recorded the percentage share of reported standstill traffic.

Figure 6 Percentage Share of Reported Standstill Traffic by Route



Weather Condition

Heavy traffic and standstill congestions most frequently occur when the weather conditions pass clouds and overcast weather and was fewest during heavy rains and rain showers probably due to suspension of work during these conditions and how infrequent these weather conditions occur.

Traffic Congestion Predictive Modeling

To foresee traffic jams, herein machine learning models were created. The data used to create the models was split up into "test" and "training" sets. The models were trained on the training dataset, and then put to the test on the testing dataset. To create this model, 18,993 data points were utilized. After partitioning the data set into a training set comprising 70% and a testing set comprising 30%, the model training was carried out. The following variables were considered in developing the models:

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Table 1 Summary of Variables used in Model Building

No	Feature	Description	Data Type
1	Traffic Congestion	Refers to the congestion level during the time of travel	Categorical
2	Speed	Refers to the speed at which the vehicle is moving	Numeric
3	Route	Refers to the specific route of travel	Categorical
4	Route Direction	Refers to the direction of route being traveled	Categorical
5	Weather Condition	Refers to the weather condition during the time of travel	Categorical
6	Day of the Week	Refers to the day of the week of travel	Categorical
7	Time	Refers to the time of the day of travel	Time
8	Month	Refers to the month of the year of travel	Categorical

When creating the model, the Variance Inflation Factor (VIF) technique is utilized to check for multi-collinearity in the independent variables. In a regression model, the VIF evaluates the correlation and correlation strength of predictor variables. The following table lists the VIF for each independent variable.

Table 2 Variance Inflation Factor Scores

Variable	VIF Score
Speed	1.250
Route	
EDSA Carousel	2.870
Quezon Avenue – EDSA Taft	1.617
Cubao – Doroteo Jose	1.216
Gilmore - Taytay	1.604
Route Direction	
North	2.405
South	1.702
East	1.310
West	1.317
Weather Condition	
Thundershowers	1.371
Light Rain	3.286
Sprinkles	1.409
Heavy Rain	1.328
Thunderstorms	1.924

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Scattered Clouds	1.678
Partly Sunny	2.779
Passing Clouds	2.458
Day	
Monday	1.271
Tuesday	1.273
Wednesday	1.609
Thursday	1.775
Friday	1.796
Saturday	1.422
Sunday	1.360
Time	1.074

The value of VIF starts at 1 and has no upper limit. A value between 1 and 5 indicates moderate correlation between a given predictor variable and other predictor variables in the model. With the above results, since none of the VIF values for the predictor variables is greater than 5, multicollinearity will not be a problem in developing the multiclass logistic regression.

Multiclass Logistic Regression

Multiclass logistic regression models represent an appropriate option when the dependent variable is categorical. This method used maximum likelihood estimation. The first iteration is a model with no regressors, only the intercept. The next iteration includes regressors in the model. The regressors were changed at every iteration, and iterations continue until the model was said to have converged.

The result of the model fitting information showed that the variables added in the model statistically and significantly improved the model compared to the intercept alone (i.e. with no variables added) with a p-value < 0.000. Hence the full model statistically and significantly predicts the dependent variable better than the intercept – only model alone. Based on the likelihood ratio tests table, as shown below all the independent variables selected are statistically significant.

Table 3 Likelihood Ratio Test

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2Log Likelihood of Reduced Model	Chi-Square	df	Sig
Intercept	47066.791a	0.000	0	0.000
Speed	60938.897	13872.107	3	0.000
Time	47191.444	124.654	3	0.000
Route	48092.261	1025.470	6	0.000
Direction	47648.206	581.415	6	0.000
Day	47923.389	856.598	18	0.000
Weather	48056.444	989.653	21	0.000
Month	47271.406	204.615	30	0.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced mode The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. *The reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom*

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

The following tables shows the important results of the multiclass logistic regression model applied on the dataset. The table on model evaluation shows that the model can accurately predict heavy and standstill traffic at 82% and 96% respectively with a precision of 76% and 71%. On the other hand, it has a recall of 85% for Heavy Traffic and 32% for Standstill Traffic. Looking at the F1-score, Heavy Traffic has 80% and 43% for the Standstill traffic. Overall, the model has a 74% accuracy with 67% F1 score at the Macro Average is used to treat all classes equally compared to the Weighted Average.

Table 4 Multiclass Logistic Regression Confusion Matrix

Actual Value	Predicted Value			
	Free Flow	Moderate	Heavy	Standstill
Free Flow	1287	443	116	5
Moderate	314	2373	713	3
Heavy	73	472	3376	66
Standstill	13	34	263	145

Table 5 Multiclass Logistic Regression Model Evaluation

	Accuracy	Precision	Recall	F1-Score
Free Flow	0.900	0.763	0.695	0.728
Moderate	0.796	0.714	0.697	0.706
Heavy	0.824	0.756	0.847	0.799
Standstill	0.960	0.662	0.319	0.430

Table 6 Multiclass Logistic Regression Overall Model Result

	Precision	Recall	F1-Score
Macro Avg	0.72	0.64	0.67
Weighted Avg	0.74	0.74	0.74

On the overall, the model has a 74% accuracy.

Random Forest

A random forest is a machine learning technique that builds decision trees on different samples and takes their majority vote for classification. It involves construction of large number of decision trees from bootstrap sample from the training dataset and selecting a subset of input features to a random subset that may be considered at each split point, it forces each decision tree to be more different.

The following table shows the results of the random forest model. The table on model evaluation shows that the random forest can predict standstill traffic at 96% accuracy with 64% precision and 53% recall which leads to 58% F1-Score. On the other hand, it exhibited a higher F1-score of 83% when it comes to predicting heavy traffic.

Table 7 Random Forest Confusion Matrix

Actual Value	Predicted Value			
	Free Flow	Moderate	Heavy	Standstill

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Free Flow	1385	351	100	15
Moderate	395	2569	421	18
Heavy	122	446	3313	106
Standstill	16	31	166	242

Table 8 Random Forest Model Evaluation

	Accuracy	Precision	Recall	F1-Score
Free Flow	0.897	0.722	0.748	0.735
Moderate	0.829	0.756	0.755	0.756
Heavy	0.860	0.828	0.831	0.830
Standstill	0.964	0.635	0.532	0.579

Table 9 Random Forest Model Overall Result

	Precision	Recall	F1-Score
Macro Avg	0.74	0.72	0.72
Weighted Avg	0.77	0.77	0.77

On the overall the model has 77% accuracy.

Neural Networks
The neural network model is constructed with 25 input attributes and one target attribute including the following parameters:

- Input layers of 25 neurons
- One hidden layer with 6 neurons
- The network has output layer of 4 neurons (i.e. four congestion levels)
- Trainlm is used as training functions
- Softmax type of activation function is used

Table 10 Neural Networks Confusion Matrix

Actual Value	Predicted Value			
	Free Flow	Moderate	Heavy	Standstill
Free Flow	3155	677	61	94
Moderate	323	2825	255	0
Heavy	35	574	1241	1
Standstill	219	44	7	185

Results showed that the model can predict standstill traffic at 96% accuracy with 66% precision and 41% recall, which results into 50% F1-Score. On the other hand, the model recorded 73% F1-Score for predicting heavy traffic.

Table 11 Neural Networks Model Evaluation

	Accuracy	Precision	Recall	F1-Score
Free Flow	0.855	0.845	0.791	0.817

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Moderate	0.807	0.686	0.830	0.751
Heavy	0.904	0.793	0.670	0.727
Standstill	0.962	0.660	0.407	0.503

Table 12 Neural Networks Overall Model Result

	Precision	Recall	F1-Score
Macro Avg	0.75	0.67	0.70
Weighted Avg	0.77	0.76	0.76

On the overall the model has 76% accuracy.

4. DISCUSSION

The accuracy rate of a model is determined by how many of its positively predicted labels are true. Positive predictive value is another name for precision. Precision and recall are used together to balance the risks of false positives and false negatives. It is important to consider the class distribution while attempting to get a precise result. Precision will suffer if there are more samples from the underrepresented group. Preciseness is a quality or accuracy metric. Selecting a model with high accuracy helps reduce the number of false negatives.
The precision score is a useful measure of the **success of prediction when the classes are very imbalanced.** Mathematically, it represents the ratio of true positive to the sum of true positive and false positive.

Table 13 Model Evaluation Result by Type of Machine Learning Model

	Accuracy	Precision	Recall	F1-Score
Multiclass Logistic Regression	0.76	0.74	0.74	0.74
Random Forest	0.77	0.77	0.77	0.77
Neural Networks	0.76	0.77	0.76	0.76

On the overall, the Random Forest shows better results compared to the two models. It has recorded the highest accuracy, recall and F1-Score. Looking at the results per class:

Table 14 Model Evaluation Result by Category and Type of Machine Learning Model

	Accuracy	Precision	Recall	F1-Score
Multiclass Logistic Regression				
Free Flow	0.900	0.763	0.695	0.728
Moderate	0.796	0.714	0.697	0.706
Heavy	0.824	0.756	0.847	0.799
Standstill	0.960	0.662	0.319	0.430
Random Forest				
Free Flow	0.897	0.722	0.748	0.735
Moderate	0.829	0.756	0.755	0.756

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

Heavy	0.860	0.828	0.831	0.830
Standstill	0.964	0.635	0.532	0.579
Neural Networks				
Free Flow	0.855	0.845	0.791	0.817
Moderate	0.807	0.686	0.830	0.751
Heavy	0.904	0.793	0.670	0.727
Standstill	0.962	0.660	0.407	0.503

Free Flow has the highest F1-score using the Neural Network, while for Moderate, Heavy, and Standstill traffic exhibited the highest F1-score using the Random Forest.

The severity of traffic congestion can be measured by the amount of time that passes during which traffic is at a standstill, as well as the number of passengers who are in the vehicle at any given time. The analysis of the predicted data model revealed that the data shows that peaks occur during those hours. In terms of the travel time gathered from the selected routes recorded takes the longest every Friday.
Weather plays a significant role in the development of traffic congestion because, in the event of flash flooding, highways become impassable to automobiles of a smaller size. Results show that there is a significant relationship between traffic congestion and weather condition data, but the Cramer V test resulted in 0.213, which indicates that there is only a weak association between the two variables. Results show that there is a significant relationship between traffic congestion and weather condition data.

5. CONCLUSION

This study aims to develop a statistical model for predicting traffic congestion. The findings of this research provide detailed traffic analysis based on the traffic data, weather conditions, and transport policies. The output of this research is to analyze factors contributing to traffic congestion and to develop a model for predicting traffic congestion by applying descriptive analysis and machine learning models.

The Random Forest model produced a higher accuracy compared to the Multinomial Logistic Regression and Neural Networks. In addition to employing information such as the route, the exact hour, and the day of the week, weather conditions were also used in the training of the model because they are significant aspects that contribute to traffic congestion.

The model's inability to achieve a higher level of accuracy can be attributed to the limited dataset. The amount of the data points needs to be increased in the future so that the accuracy of the model may be enhanced. In addition, both real-time and historical data must be integrated in order to achieve an even better precision level. Moreover, the prediction models can be further improved by considering road condition and road crash data.

The results of this research have the potential to be incorporated into the design of a dashboard that will be used for the monitoring of traffic. The findings of this study can be utilized by

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 03, No. 04, June-July 2023
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.34.13.28

national transportation organizations in the process of creating, developing, and putting into operation the dashboard to monitor high-traffic regions and devise plans for the regions' overall improvement.

6. REFERENCES

1. Dato V, et al. (2010). Philippines Discussion Note No. 8: Transport for Growth and Integration. Manila: The World Bank
2. Department of Transportation. Metro Manila Traffic Watch. Retrieve from https://sites.google.com/view/metromanilatrafficwatch/metro-manila-traffic-watch
3. Department of Transportation. Sectoral Data Sets. Retrieved from https://dotr.gov.ph/data-sets.html
4. Deshpande M. and Bajaj P. (2017). Performance improvement of traffic flow prediction model using combination of support vector machine and rough set. International Journal of Computer Applications, vol. 163, no. 2, pp. 31–35, 2017
5. Japan International Cooperation Agency. (2019). Follow-up Survey on Roadmap for Transport Infrastructure Development for Greater Capital Region (GCR) Final Report Summary. Retrieved from https://openjicareport.jica.go.jp/pdf/1000041638_03.pdf
6. Land Transportation Office. (2021). Annual Report. Retrieved from https://lto.gov.ph/transparency-seal/annual-reports.html
7. Ochave, R. M. (2022). PCCI seeks immediate action on NCR traffic to remove drag on economy. Retrieved from https://www.bworldonline.com/economy/2022/05/09/447325/pcci-seeks-immediate-action-on-ncr-traffic-to-remove-drag-on-economy/
8. Santos, L. (2020). Influence of Traffic Congestion in Business Development: A Literature Review
9. TomTom Traffic Index. (2019). Manila in Traffic Index. Retrieved from https://www.tomtom.com/traffic-index/manila-traffic/