July – August 2020
ISSN: 0193-4120 Page No. 390 - 394

A Novel Framework for Traffic Management using Machine Learning

Deepak Dudeja
Ph.D. Research Scholar, DCSA
CT University, Ludhiana
9466794142, deepakdudeja1983@gmail.com

Harmeet Singh
Assistant Professor,DCSA
CT University, Ludhian
09877002246, harmeet17333@ctuniversity.in

Prashant Singh Rana
ASSISTANT PROFESSOR, TIET, PATIALA
09313889932, psrana@gmail.com

Article Info
Volume 83
Page Number: 390 - 394
Publication Issue:
July - August 2020

Article History
Article Received*: 06 June 2020*
Revised*: 29 June 2020*
Accepted*: 14 July 2020*
Publication*: 25 July 2020*

Abstract
Traffic Congestion is becoming a worldwide problem, it is necessary to find out the solution. It is necessary to find Traffic Congestion factors for reducing the level of the congestion. Addressing Traffic Congestion is a challenging and time-demanding task that requires a large research study to ensure successful observing situations. Our systematic approach uses modeling on travel situations for the purpose of analyzing the state of traffic and offers an immediate solution to travelers and traffic administrators. This is possible by deploying to measure the level of congestion at a particular time and its magnitude. By having prior knowledge of traffic congestion, the traveler can take a decision. The significance of modeling helps in saving cost, time and money. Therefore, it is necessary to develop foolproof solutions by creating new modeling ideas or combining existing modeling techniques to address the issues of road safety support, human life safety support, and reducing traffic congestion. This paper discusses about modeling the level of traffic congestion and gives the best solution.

Keywords— CM, DFD, LR, SVM, GPS

I. INTRODUCTION

The traffic management system has obstacles like shortage of budget, un-planned cities, non-discipline, and out-of-date management techniques. Road traffic congestion is increasing in
the world, and its impact is visible in the form of fuel and time wastage. The poorly planned network is the most common issue behind traffic congestion. Traffic congestion problem happens due to several factors like slow drive due to traffic delay (Timalsena, Marsani, and Tiwari, 2017), crowd at the tourist places (Iyer, Boxer and Subramanian, 2018) , Natural disasters (Hara and Kuwahara, 2015) , Improper Planning (Lal et al., 2016) , Construction work on road (Hyari, El-Mashaleh and Rababeh, 2015) and unplanned events (Humphreys and Pyun, 2018). Some of the phenomenon related to traffic congestion are stress (Vencataya et al., 2018), speeding (Chang et al., 2019), death (Organization, 2015), frustration (Hickman et al., 2018), time wastage (Raheem et al., 2015), pollution (Li and Huang, 2019) and low public satisfaction (Nguyen, 2019). Travel behavior has a direct impact on these factors, and it is one of the most important theories used to cater to the problem of traffic congestion and management.
The structure of the paper in different sections, such as section 1. Data Collection 2. Methodology 3. Model Implementation 4. provides validation of the model, 5. Conclusion.

II. DATA COLLECTION

The downloaded dataset travel times from web and the brief about dataset is given below:

July – August 2020
ISSN: 0193-4120 Page No. 390 - 394

The driver used an application to sign GPS coordinates as he goes to work and back daily through driving. The application gets the location and elevation data. We have total 213 records in the dataset.
We have 12 variables in the dataset as given below:

1. Date of travel: It tells the date on which driver went.
2. StartTime: Indicates the times when driver starts the vehicle
3. DayOfWeek: Shows the day on which he went.
4. GoingTo: Means the direction of travel.
5. Distance : It shows the distance travelled in kilometers
6. MaxSpeed: It highlights fastest speed recorded (all trips are on the 407 highway for some portion)
7. AvgSpeed: It indicates the average speed for the entire trip
8. AvgMovingSpeed: It tells the average speed recorded only while the car is moving
9. FuelEconomy: It highlights a rough estimate of fuel economy.
10. TotalTime: It gives the total time of the entire trip, in minutes
11. MovingTime: It tells duration when the car was considered to be moving (i.e. not counting traffic delays, accidents, or time while the car is stationary)
12. Take407All: It is Yes if the 407 toll highway was taken for the entire trip. I try to avoid taking the 407, taking slower back routes to save costs. But some days I'm running late, or just lazy, and take it all the way.

First, We have implemented data exploratory functions like dim, sum, str etc. and check whether there is any missing values or outliers. When we analyse the data we can see that Take407all is the dependent variable and all others are independent variable. We can also calculate the density of Distance as per figure 1.1

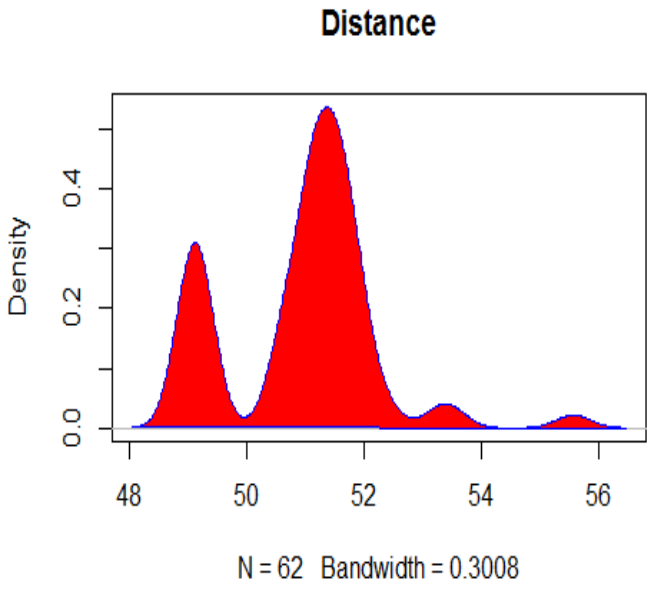

Figure 1.1 : Density of Distance

Relationship between Distance and Speed
We can represent the relationship between Distance and Speed using boxplot graphs as Follows:

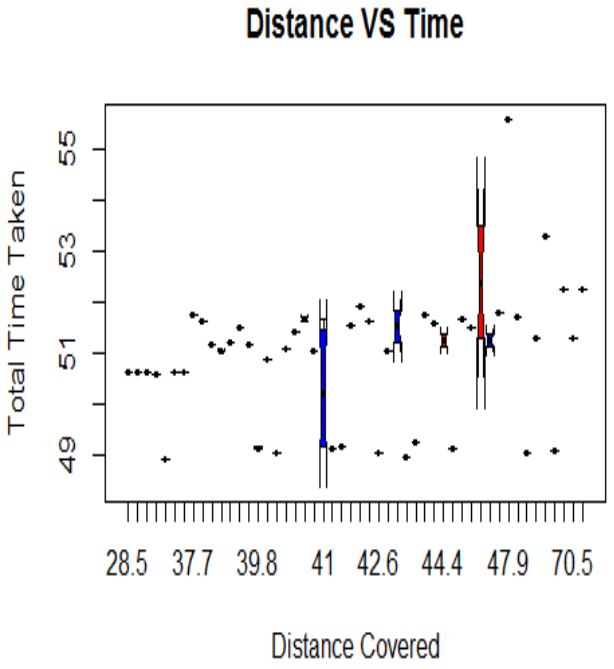

Figure 1.2: Relationship between Distance and Time

July – August 2020
ISSN: 0193-4120 Page No. 390 - 394

Data Flow Diagram (DFD): DFD of TRAVEL MODEL IS AS SHOWN IN FIG. 1.3

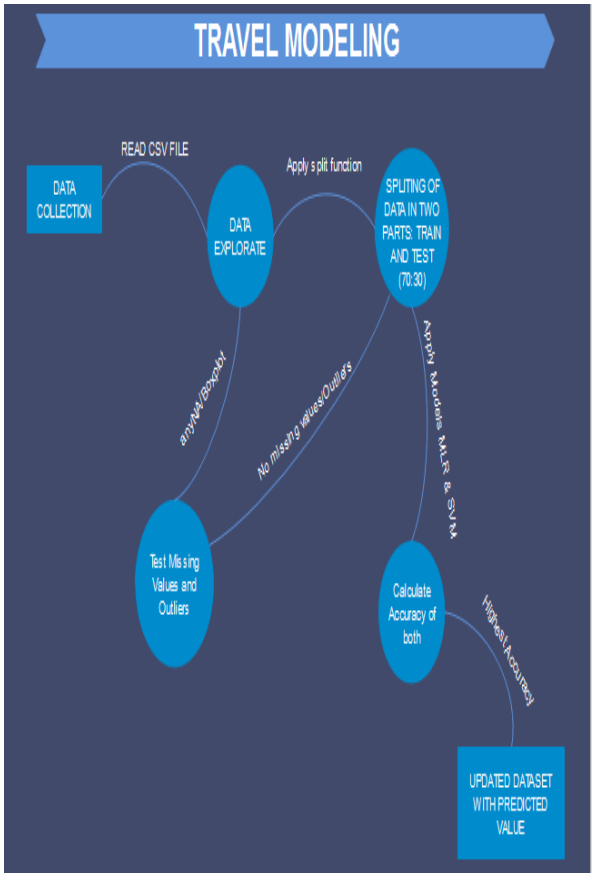

Figure 1.3 : DFD of Travel Modeling

3. Methodology

We have used two methods to solve the problem of traffic :

1. Logistic Regression Model
2. SVM using (Linear Kernel and Sigmoid Kernel)

3.1 Logistic Regression Model

Logistic Regression model is widely used for binary classification problems. It can also be extended to multi-class classification problems. Here the dependent variable is categorical i.e. $y \in \{0,1\}$.
Binary dependent value can have only two values 0 or 1. In this paper, binary variable is Take407All.

$$P(y = 1 \mid x; \theta) = h_\theta(x) \quad \ldots\ldots\ldots\ldots(1)$$

$$P(y = 0 \mid x; \theta) = 1 - h_\theta(x) \quad \ldots\ldots\ldots\ldots\ldots(2)$$

First equation shows the probability of output variable y being equal to 1 i.e. P (y=1). It is equal to sigmoid (σ) of θT_x.
Note that θT_x is vector notation for

$$\theta 1x1 + \theta 2x2 + \ldots\ldots\ldots + \theta nxn.$$

The second equation shows probability of output variable y being equal to zero P(y)=0.
The total of two probabilities is 1.

3.2 Support Vector Machine

Support Vector Machines (SVM) are classification algorithms used to assign data to various classes. They involve detecting hyperplane which segregate into classes. The hyperplane chosen should be such that all the data points in the dataset are far away from it.

Any hyperplane can be written as the set of points x satisfying

$$w.x - b = 0$$

where w is the (not necessarily normalized) normal vector to the hyperplane. This is much like Hesse normal form, except that w is not necessarily a unit vector.

SVM can not only perform linear classification but also non linear classification using what is called the kernel trick. For linear classification, we have used linear kernel and in non linear we have used sigmoid kernel.

3.3 Technique for Validating the Model : Confusion Matrix (CM)

A confusion matrix is a method to encapsulate the performance of a classification algorithm.

Classification accuracy is useless if you have an unequal number of observations in each class or if you have more than two classes in your dataset. Calculating a confusion matrix can give you a better idea of what your classification model is getting exact and what types of errors it is making. Confusion matrix generate a matrix like this, then we can calculate accuracy from this matrix

Then we can calculate accuracy using the formula (A11+A22)/(A11+A12+A21+A22). In this example, accuracy is 95%

4. Model Implementation:

For this dataset, we have first apply Logistic Regression Model, for that we have to split the dataset in two parts in the ratio of 70:30 : train dataset and test dataset respectively. We apply the on train dataset:

July – August 2020
ISSN: 0193-4120 Page No. 390 - 394

regmod<-glm(Take407All~.,data = train_data,family=binomial())

After that we get the significant variables (value of p > 0.0.5). In this model we got three significant variables Distance, AvgMovingSpeed and total time.
Once again we applied model on these three significant variables:

final_model<-glm(Take407All~Distance+AvgMovingSpeed+TotalTime, data= train_data,family=binomial())

then we calculate the predicted values through test data and we get the final predicted values. We have bind these values in the existing dataset using cbind function.
In second case, we implemented Support Vector Machine, same steps performed like splitting then apply the model svm on the given dataset:

classifier<- svm(Take407All ~ ., data= train_data, type='C-classification', kernel='linear')

When we use SVM using linear kernel then number of support vectors are 25 and in case of SVM using Sigmoid kernel then number of support vectors are 38.
classifier1<- svm(Take407All ~ ., data=train_data, type='C-classification', kernel='sigmoid')

5. Validations of Model

We have implemented two different machine learning models: Logistic Regression Model, Support Vector Machine. After Implementation, validate the accuracy through Confusion Matrix. It is found that the accuracy of Logistic Regression is high as compared to others.

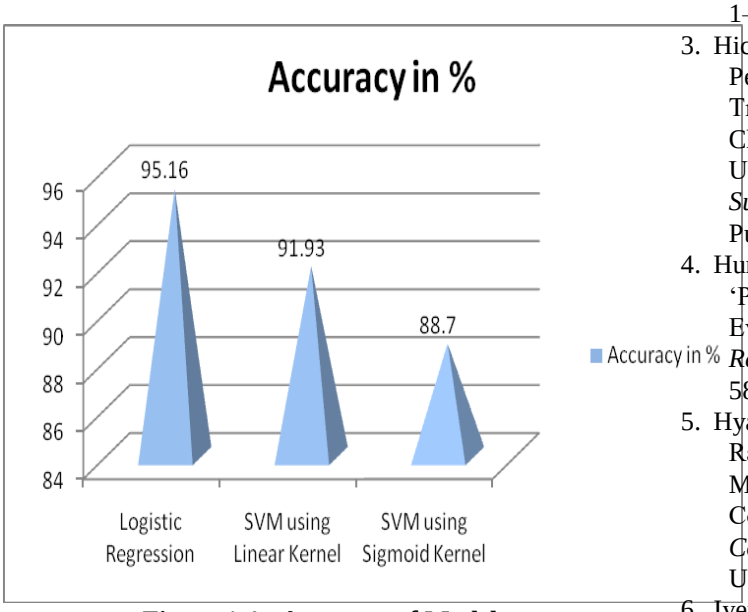

Figure 1.4 : Accuracy of Models

6. Conclusion

Modeling the traffic situation is becoming the challenging task because of not having accurate model of level of traffic on the road. In our dataset, earlier if user takes GT Road (Take407All) then Total time will be 28 minutes to 38 minutes but when he travel through side by then total time will be 34 minutes to 61 minutes. The proposed model calculated the predicted variable that tells if time of travel will be more than 30 minutes then user faced the traffic in case of Take407All. In other case, if time of travel was more than 35 minutes then user also faced traffic. The proposed model addresses the problem of traffic using two different models. The two models were involved such as Logistic Regression Model, Support Vector Machine Model The percentage accuracy level of Logistic Regression model is 95.16%, SVMM using Linear Kernel is 91.93% and SVMM with Sigmoid Kernel is 88.7% using confusion matrix. The proposed model shows about the two paths such as toll based path and without toll based. The traveler can choose the path according to current traffic siltation. The proposed method is the showing the alternative path for consuming time and cost. The can work in the future for getting more efficiency of the model.

References

1. Chang, X. *et al.* (2019) 'Effects of on-Board Unit on Driving Behavior in Connected Vehicle Traffic Flow', *Journal of Advanced Transportation*. Hindawi, 2019.
2. Hara, Y. and Kuwahara, M. (2015) 'Traffic Monitoring immediately after a major natural disaster as revealed by probe data–A case in Ishinomaki after the Great East Japan Earthquake', *Transportation research part A: policy and practice*. Elsevier, 75, pp. 1–15.
3. Hickman, R. *et al.* (2018) '"I Drive outside of Peak Time to Avoid Traffic Jams—Public Transport Is Not Attractive Here." Challenging Discourses on Travel to the University Campus in Manila', *Sustainability*. Multidisciplinary Digital Publishing Institute, 10(5), p. 1462.
4. Humphreys, B. R. and Pyun, H. (2018) 'Professional sporting events and traffic: Evidence from US cities', *Journal of Regional Science*. Wiley Online Library, 58(5), pp. 869–886.
5. Hyari, K. H., El-Mashaleh, M. S. and Rababeh, S. M. (2015) 'Framework for Managing the Traffic Impacts of Building Construction Projects', *Journal of Construction in Developing Countries*. Universiti Sains Malaysia Press, 20(2), p. 97.
6. Iyer, S. R., Boxer, K. and Subramanian, L.

July – August 2020
ISSN: 0193-4120 Page No. 390 - 394

(2018) 'Urban Traffic Congestion Mapping Using Bus Mobility Data.', in *UMCit@ KDD*, pp. 7–13.

7. Lal, G. *et al.* (2016) 'Sustainable traffic improvement for urban road intersections of developing countries: a case study of Ettumanoor, India', *Procedia technology*. Elsevier, 25, pp. 115–121.
8. Li, Z. and Huang, J. (2019) 'How to Mitigate Traffic Congestion Based on Improved Ant Colony Algorithm: A Case Study of a Congested Old Area of a Metropolis', *Sustainability*. Multidisciplinary Digital Publishing Institute, 11(4), p. 1140.
9. Nguyen, X. P. (2019) 'The Bus Transportation Issue And People Satisfaction With Public Transport In Ho Chi Minh City', *Journal of Mechanical Engineering Research & Developments (JMERD)*. Zibeline International Publishing, 42(1), pp. 10–16.
10. Organization, W. H. (2015) *Global status report on road safety 2015*. World Health Organization.
11. Raheem, S. B. *et al.* (2015) 'The Cause, Effect and Possible Solution to Traffic Congestion on Nigeria Road (A Case Study of Basorun-Akobo Road, Oyo State)', *International Journal of Engineering Science Invention*, 4(9), pp. 10–14.
12. Timalsena, A. P., Marsani, A. and Tiwari, H. (no date) 'Impact of Traffic Bottleneck on Urban Road: A Case Study of Maitighar–Tinkune Road Section'.
13. Vencataya, L. *et al.* (2018) 'Assessing the Causes & Impacts of Traffic Congestion on the Society, Economy and Individual: A Case of Mauritius as an Emerging Economy', *Studies in Business and Economics*. Sciendo, 13(3), pp. 230–242.
14. Retrieved from https://openmv.net/ and https://machinelearningmastery.com/confusion-matrix-machine-learning/

Author Profile

Deepak Dudeja, Google And Ibm Certified, 12 Years Of Teaching Experience, Published Two Books (Artificial Intelligence And Software Engineering) And 12 Research Papers, Good Knowledge Of Data Science And Machine Learning And Done Projects On The Same. Currently, Ph.D. Research Scholar at CT University, Ludhiana, Punjab.

Dr. Prashant Singh Rana, Assistant Professor at TIET, Patiala, Punjab. He has 31 SCI and 22 Conference Publications. His Area of Specialization is Machine Learning. He has 7 patents and taken grant for 5 projects from Government.

Dr. Harmeet Singh, Assistant Professor, CT University, Ludhiana. He has done Ph.D. in Biorobotics at Scuola Superiore Sant'Anna. He is also project head of Robonova Start Up at CT University