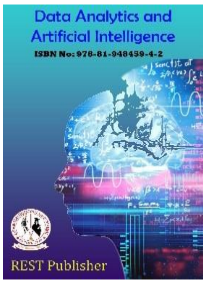
Data Analytics and Artificial Intelligence
Vol: 3(7), 2023
REST Publisher; ISBN: 978-81-948459-4-2
Website: http://restpublisher.com/book-series/daai/



Traffic Prediction Using Machine Learning

M. Mohan Raj
Adhiyamaan College of Engineering Hosur, Tamilnadu, India.
*Corresponding Author Email: @mohanraj.mca2023@adhiyamaan.in

Abstract: *The paper deals with traffic prediction that can be done in intelligent transportation systems which involve the prediction between previous year's data set and he recent year data which ultimately provides the accuracy and mean square error. This prediction will be helpful for the people who are in need to check the immediate traffic state. The traffic data is predicated on a basis of 1 hour time gap. Live statistics of the traffic is analyzed from this prediction. So this will be easier to analyze when the user is on driving too. The system compares the data of all roads and determines the most populated roads of the city. I propose the regression model in order to predict the traffic using machine learning by importing Sklearn, Keras and Tensorflow libraries*
Keywords: *Traffic, Regression, Intelligent Transport System (ITS), Machine learning, Prediction*

1. INTRODUCTION

Machine Learning (ML) is one of the most important and popular emerging branches these days as it are a part of Artificial Intelligence (AI). This is possible only when the traffic flow is smooth. To deal with this, Traffic prediction is needed so that we can estimate or predict the future traffic to some extent. In recent times, machine learning becomes an essential and upcoming research area for transportation engineering, especially in traffic prediction. Traffic congestion affects the country's economy directly or indirectly by its means. Traffic congestion also takes people's valuable time, cost of fuel every single day. As traffic congestion is a major problem for all classes in society, there has to be a small-scale traffic prediction for the people's sake of living their lives without frustration or tension. For ensuring the country's economic growth, the road user's ease is required in the first place. Introduction, Purpose of Traffic Prediction, Problem Statement, Related Work, Overview, Methodology, Software Implementation and Conclusion with Future work.

2. RESEARCH OBJECTIVES AND GOAL

The research area is on Traffic Flow Prediction, focusing on urban road Traffic flow prediction using ML techniques. Many existing traffic flow prediction models are unable to give accurate results on predicting traffic jams. This weakness has sparked interest in developing a ML model for traffic flow prediction. ML methods can be quite versatile and effective in making this model. A literature review has been undertaken to understand ITS as well as Traffic Jam Prediction (TJP) and how ML can be used for TFP. The role of ML in TJP is also examined. The objectives of this study are:

- How will emerging technologies, such as ML, improve the ITS?
- The importance of ITS in urban cities.
- Importance of urban road TFP Model.
- What kinds of the dataset are required for a TFP system? To identify research gaps and attempt to provide a remedy

2.1 methodologies followed: Various journal databases such as Elsevier, Science Direct, IEEE, Google scholar, and others have been used for this purpose and shortlisted articles that used ML, Deep Learning, and ANN techniques to predict the urban road traffic jam.All study materials were collected initially, and studied to find key threads across the articles.

2.2. Overview of tjp in ml: In this section, using ML in ITS, with a particular emphasis on how ML increase perception, prediction, and management duties, among other things is addressed. Introduction to the ML methodologies, including terminology and ideas that are commonly found in the literature, are provided.

3. OVERVIEW

In traffic congestion forecasting there are data collection and prediction model.

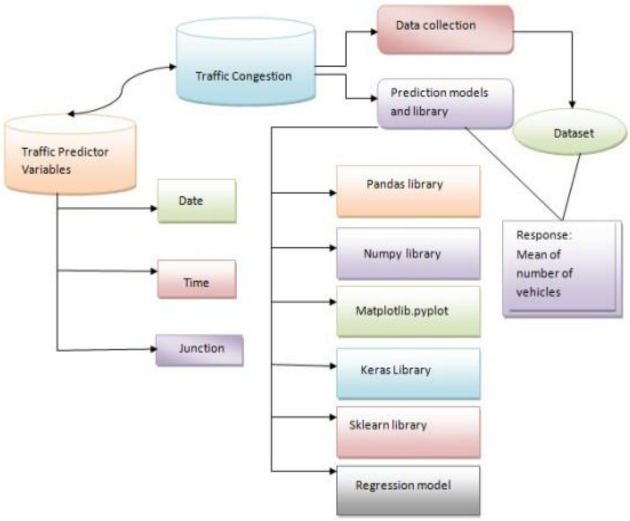


FIGURE 1. Outline of the traffic prediction in this paper

The methodology has to be done correctly so that there won't be any flaws while predicting. After data collection, the vital role is the data processing which is to train and test the datasets that is taken as the input. After processing the data, the validation of the model is done by using necessary models. Figure 1 highlights the outline of traffic prediction using machine learning.

4. MACHINE LEARNING PREDICTIONS IN ITS

ML techniques have achieved a high level of performance on prediction challenges in ITS, primarily delivering tasks that may be classified into predicting traffic flow, travel time, vehicle behaviour, user behaviour, and road occupancy [10][11]. Table 1 shows

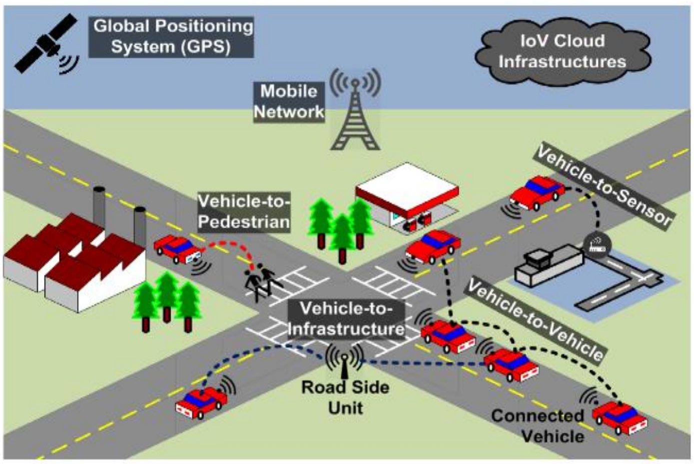


FIGURE 2. Machine Learning Predictions in Its

4.1. Methodology: Many researchers have been used various discussed approaches. This paper contains the technique of predicting the traffic using regression model using various libraries like Pandas, Numpy, OS, Matplotlib. pyplot, Keras and Sklearn.

4.1.2. Data set: Traffic congestion is raising a lot these days. Factors like expanding urban populations, uncoordinated traffic signal timing and a lack of real-time data. The effect of the traffic congestion is very huge these days. Data collected in this paper are from the Kaggle website for the implementations of machine learning algorithms using python3 to show outputs in the traffic prediction. Two datasets are collected in which one is the 2015's traffic data which comprises of date, time, and number of vehicles, junction and the rest one is the 2017's traffic data with the same details so as to compare easily without any misconception. The unwanted data has been deleted by pre-processing the data aggregated from 1 to 24 hours time interval to calculate traffic flow prediction with each 1 hour interval. [11]

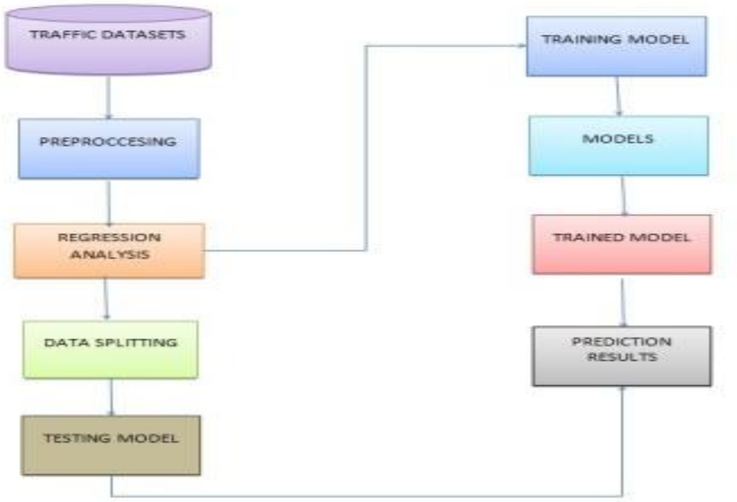


FIGURE 2. Regression model of traffic prediction in this paper

The evaluation yields a foretold value for the benchmark resulting from a sum of scalar vectors of the predictors. The accuracy is measured by computing mean square error. Thus obtaining the expected error from the observed value and also truth value which is equivalent to the standard deviation deployed within the statistical methods [4],[12]. Figure 2 shows the Regression model for the Traffic Prediction.

5. SOFTWARE IMPLEMENTATION

5.1. Simulation: The evaluation yields a foretold value for the benchmark resulting from a sum of scalar vectors of the predictors. The accuracy is measured by computing mean the command prompt is the local host in this paper to initialize the Jupiter book. The local host contains the extensions which we modify to our convenience.

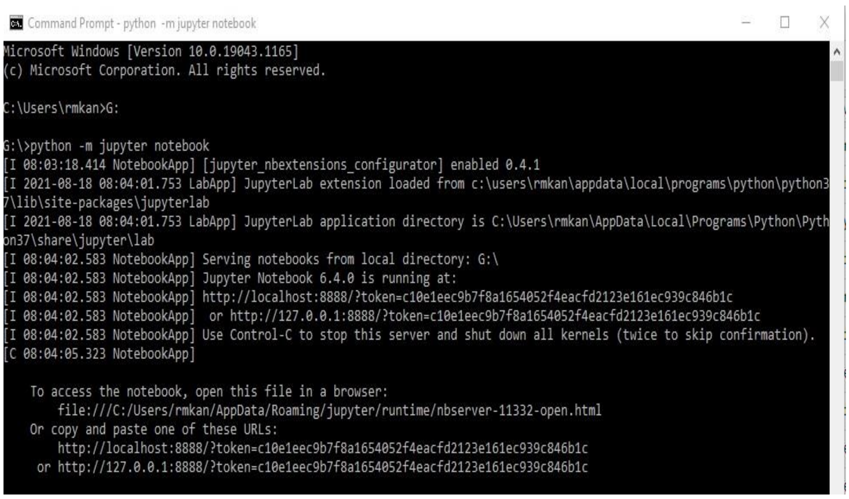


FIGURE 3. The figure signifies the initializing the jupyter notebook through command prompt

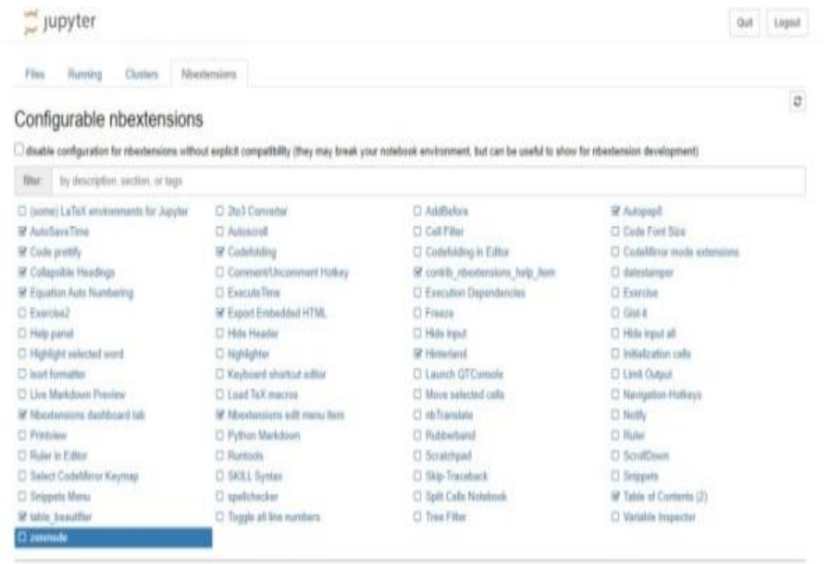


FIGURE 4. The figure denotes the necessary extensions that is needed for the prediction to take place

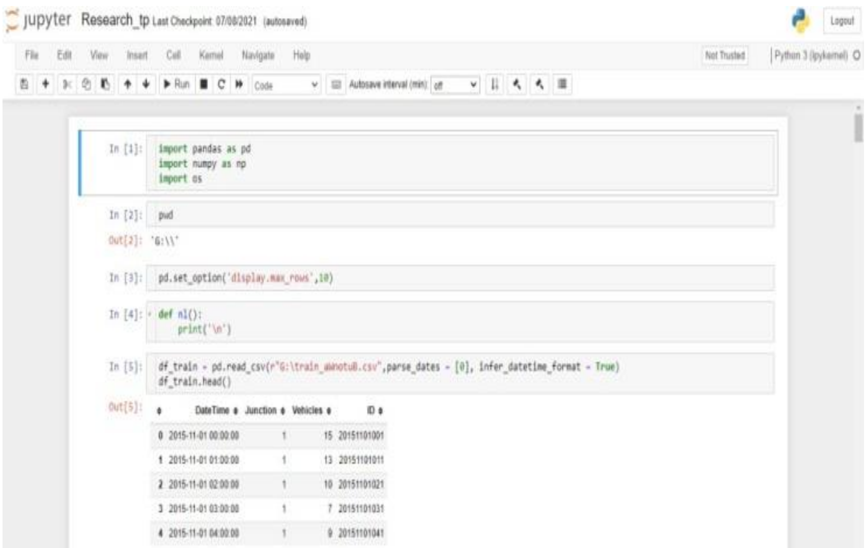


FIGURE 5. The figure implies the page of the jupyter notebook

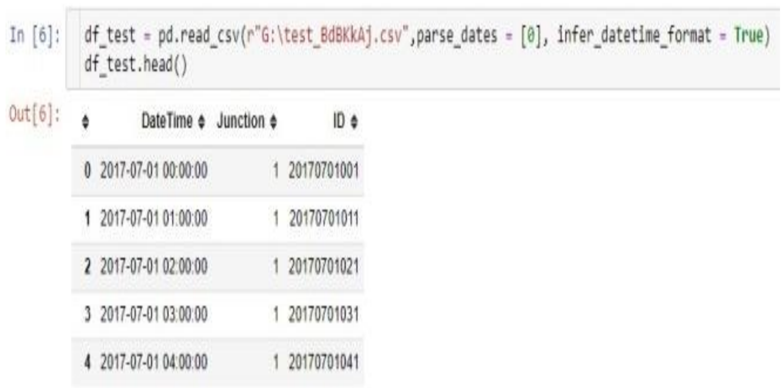


FIGURE 6. The figure above denotes the result after running the particular command in the prediction process will be displayed under out command

The below mentioned are the libraries used for the prediction of traffic import pandas as pd # library imported import numpy as np # library imported import os # library imported

TABLE 1. The training model of the junction 1 for the prediction that is taken[19]

Date Time	Junction	Vehicles	ID
0 2015-11-01 00:00:00	1	15	20151101001
1 2015-11-01 01:00:00	1	13	20151101011
2 2015-11-01 02:00:00	1	10	20151101021
3 2015-11-01 03:00:00	1	7	20151101031
4 2015-11-01 04:00:00	1	9	20151101041

TABLE 2. The testing model of the junction 1 for the prediction that is taken

Date Time	**Junction**	**ID**
2017-07-01 00:00:00	1	20170701001
2017-07-01 01:00:00	1	20170701011
2017-07-01 02:00:00	1	20170701021
2017-07-01 03:00:00	1	20170701031

The training model and testing model of Junction 1 is displayed using the following commands

time_targets = df_tmp.groupby([level_values(0)] + [pd. Grouper(freq = '1M',level= -1)])['Vehicles'].sum()

train = df_train.pivot(index = 'DateTime',columns = 'Junction', values = 'Vehicles') train = train.fillna(0)

TABLE 3. The pivot table of both the training and testing model without NaN values

Junction	**1**	**2**	**3**	**4**
DateTime				
2015-11-01 00:00:00	15.0	6.0	9.0	0.0
2015-11-01 01:00:00	13.0	6.0	7.0	0.0
2015-11-01 02:00:00	10.0	5.0	5.0	0.0
2015-11-01 03:00:00	7.0	6.0	1.0	0.0
2015-11-01 04:00:00	9.0	7.0	2.0	0.0
...	...	...	...	...
2017-06-30 19:00:00	105.0	34.0	33.0	11.0
2017-06-30 20:00:00	96.0	35.0	31.0	30.0
2017-06-30 21:00:00	90.0	31.0	28.0	16.0
2017-06-30 22:00:00	84.0	29.0	26.0	22.0

The training model and testing model of Junction 1 is displayed using the following commands

Xy_train =gen_lag_features(train)

TABLE 4. The table of both the training and testing model of concatenated values

Junction 1(H-1)	Junction 2(H-1)	Junction 3(H-1)	Junction 4(H-1)	Junction 1(H)	Junction 2(H)	Junction 3(H)	Junction 4(H)
DateTime							
2015-11-01 01:00:00	15.0	6.0	9.0	0.0	13.0	6.0	7.0
2015-11-01 02:00:00	13.0	6.0	7.0	0.0	10.0	5.0	5.0
2015-11-01 03:00:00	10.0	5.0	5.0	0.0	7.0	6.0	1.0
2015-11-01 04:00:00	7.0	6.0	1.0	0.0	9.0	7.0	2.0
2015-11-01 05:00:00	9.0	7.0	2.0	0.0	6.0	2.0	2.0
...	...	...	...	...	...	...	...
2017-06-30 19:00:00	95.0	34.0	38.0	17.0	105.0	34.0	33.0
2017-06-30 20:00:00	105.0	34.0	33.0	11.0	96.0	35.0	31.0
2017-06-30 21:00:00	96.0	35.0	31.0	30.0	90.0	31.0	28.0
2017-06-30 22:00:00	90.0	31.0	28.0	16.0	84.0	29.0	26.0
2017-06-30 23:00:00	84.0	29.0	26.0	22.0	78.0	27.0	39.0

Command here used:

Xy_train[Xy_train.columns]=scaler.fit_transform(Xy_train[Xy_train.columns])

TABLE 5. The table contains the training model's results where the scaling of the training data is simulated

Junction 1(H-1)	Junction 2(H-1)	Junction 3(H-1)	Junction 4(H-1)	Junction 1(H)	Junction 2(H)	Junction 3(H)	Junction 4(H)
DateTime							

2015-11-01 01:00:00	0.066225	0.106383	0.044693	0.000000	0.052980	0.106383	0.033520
2015-11-01 02:00:00	0.052980	0.106383	0.033520	0.000000	0.033113	0.085106	0.022346
2015-11-01 03:00:00	0.033113	0.085106	0.022346	0.000000	0.013245	0.106383	0.000000
2015-11-01 04:00:00	0.013245	0.106383	0.000000	0.000000	0.026490	0.127660	0.005587
2015-11-01 05:00:00	0.026490	0.127660	0.005587	0.000000	0.006623	0.021277	0.005587
...	...	...	...	...	...	...	...
2017-06-30 22:00:00	0.562914	0.638298	0.150838	0.444444	0.523179	0.595745	0.139665
2017-06-30 23:00:00	0.523179	0.595745	0.139665	0.611111	0.483444	0.553191	0.212291

Command here used: X_train = Xy_train[Xy_train.index< '2017-04-01'].iloc [:,0:4]

TABLE 6. The table contains the training model's first 4 rows and columns and the testing model's first 4 rows and columns[14]

Junction 1(H-1)	Junction 2(H-1)	Junction 3(H-1)	Junction 4(H-1)
DateTime			
2015-11-01 01:00:00	0.066225	0.106383	0.044693
2015-11-01 02:00:00	0.052980	0.106383	0.033520
2015-11-01 03:00:00	0.033113	0.085106	0.022346
2015-11-01 04:00:00	0.013245	0.106383	0.000000
2015-11-01 05:00:00	0.026490	0.127660	0.005587
...	...	...	...
2017-03-31 19:00:00	0.476821	0.574468	0.178771
2017-03-31 20:00:00	0.496689	0.531915	0.156425
2017-03-31 21:00:00	0.483444	0.638298	0.156425
2017-03-31 22:00:00	0.403974	0.574468	0.150838
2017-03-31 23:00:00	0.423841	0.553191	0.162011

Command here used:y_train = Xy_train[Xy_train.index< '2017-04-01'].iloc [:,4:] y_train me = pd.concat([d1,d2],axis = 1,join = 'outer')

TABLE 7. The table contains the combining of the two datasets

	DateTime	Junction	Vehicles	ID	DateTime	Junction	ID
0	2015-11-01 00:00:00	1	15	20151101001	2017-07-01 00:00:00	1	20170701001
1	2015-11-01 01:00:00	1	13	20151101011	2017-07-01 01:00:00	1	20170701011
2	2015-11-01 02:00:00	1	10	20151101021	2017-07-01 02:00:00	1	20170701021
3	2015-11-01 03:00:00	1	7	20151101031	2017-07-01 03:00:00	1	20170701031
4	2015-11-01 04:00:00	1	9	20151101041	2017-07-01 04:00:00	1	20170701041
...	...	...	...	...	...	...	...
11806	2017-03-06 22:00:00	1	81	20170306221	2017-10-31 22:00:00	4	20171031224
11807	2017-03-06 23:00:00	1	72	20170306231	2017-10-31 23:00:00	4	20171031234

5.2. Results from the simulation:

The results of the traffic are as follows which by the matplotlib library[15].

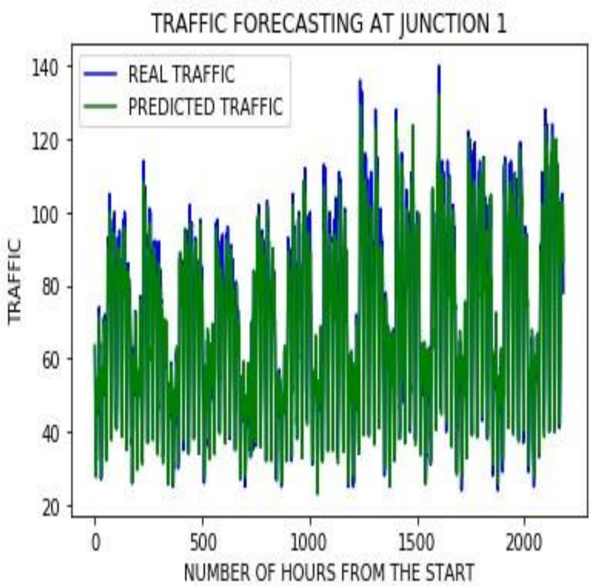


FIGURE 7. The figure signifies the Traffic prediction of Junction 1 from the datasets

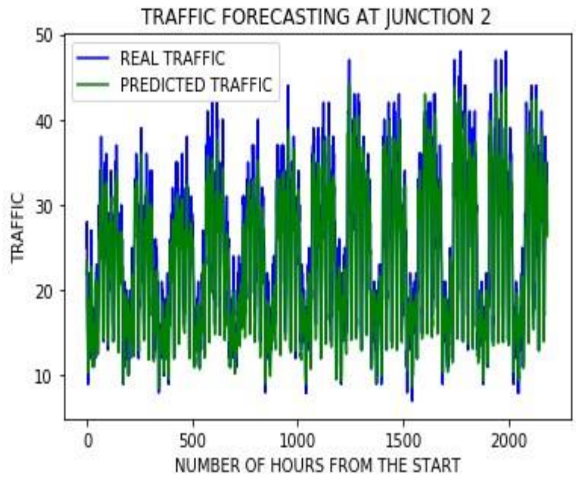


FIGURE 8. The figure signifies the Traffic prediction of Junction 2 from the datasets

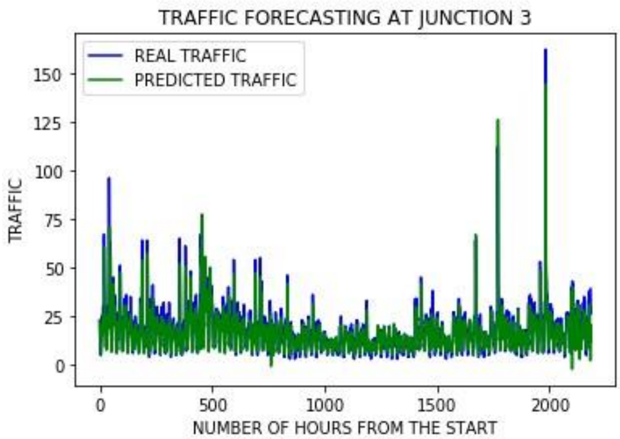


FIGURE 9. The figure signifies the Traffic prediction of Junction 3 from the datasets

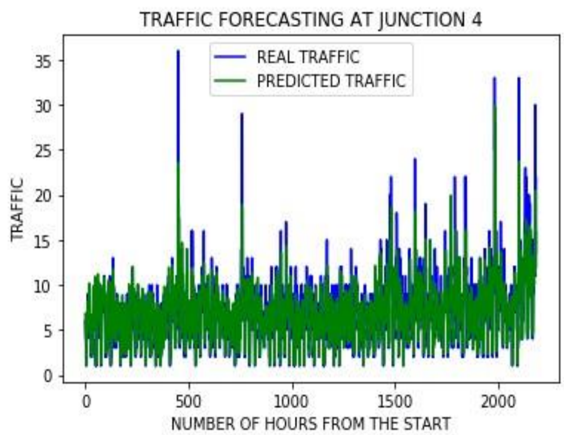


FIGURE 10. The figure signifies the Traffic prediction of Junction 4 from the datasets

İn the last, researchers are suggetsed to go through various uses of Machine Learning in many useful applications like churn prediction, healthcare, agriculture, Transportation, etc., in [21-32].

6. CONCLUSION

In the system, it has been concluded that we develop the traffic flow prediction system by using a machine learning algorithm. By using regression model, the prediction is done. The public gets the benefits such as the current situation of the traffic flow, they can also check what will be the flow of traffic on the right after one hour of the situation and they can also know how the roads are as they can know mean of the vehicles passing though a particular junction that is 4 here. The weather conditions have been changing from years to years. The cost of fuel is also playing a major role in the transportation system. Many people are not able to afford the vehicle because of the fuel cost. So, there can be many variations in the traffic data. There is one more scenario where people prefer going on their own vehicle without carpooling, this also matters in the traffic congestion. So, this prediction can help judging the traffic flow by comparing them with these 2 years data sets. The forecasting or the prediction can help people or the users in judging the road traffic easier before hand and even they can decide which way to go using their navigator and also this will prediction will be also helpful.

7. FUTURE WORK

In the future, the system are often further improved using more factors that affect traffic management using other methods like deep learning, artificial neural network, and even big data. The users can then use this technique to seek out which route would be easiest to achieve on destination. The system can help in suggesting the users with their choice of search and also it can help to find the simplest choice where traffic isn't in any crowded environment. Many forecasting methods have already been applied in road traffic jam forecasting. While there's more scope to create the congestion prediction more precise, there are more methods that give precise and accurate results from the prediction. Also, during this period, the employment of the increased available traffic data by applying the newly developed forecasting models can improve the prediction accuracy. These days, traffic prediction is extremely necessary for pretty much every a part of the state and also worldwide. So, this method of prediction would be helpful in predicting the traffic before and beforehand. For better congestion prediction, the grade and accuracy are prominent in traffic prediction. within the future, the expectation are going to be the estimation of established order accuracy prediction with much easier and user-friendly methods so people would find the prediction model useful and that they won't be wasting their time and energy to predict the information. There will be some more accessibility like weather outlook, GPS that's the road and accident-prone areas will be highlighted in order that people wouldn't prefer using the paths which aren't safe and simultaneously they'll predict the traffic. This will be done by deep learning, big data, and artificial neural networks.

REFERENCES

[1]. Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Optimized and Meta-Optimized Neural Networks for Short Term Traffic Flow Prediction: A Genetic Approach. Transportation Research Part C: Emerging Technologies, Vol. 13, No. 3, 2005, pp. 211–234.
[2]. Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment AmrElfar1, Alireza Talebpour2, and Hani S. Mahmassani1, National Academy of Sciences: Transportation Research Board 2018.
[3]. Big data-driven machine learning-enabled traffic flow prediction Fanhui Kong1 Jian Li1 Bin Jiang2 Tianyuan Zhang3 Houbing Song3, 2018.
[4]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2845248/
[5]. https://jupyter.org/
[6]. Bao G, Zeng Z, Shen Y. Region stability analysis and tracking control of memristive recurrent neural network. Neural Netw. 2017;5(1):74-89.
[7]. Jiang, X., and H. Adeli.Dynamic Wavelet Neural Network Model for Traffic Flow
[8]. Forecasting. Journal of Transportation Engineering, Vol. 131, No. 10, 2005, pp. 771– 779
[9]. Identification Traffic Flow Prediction Parameters Anuchit Ratanaparadorn Department of Industrial Engineering, Kasetsart University, Thailand, Anuchit Ratanaparadorn, Sasivimol Meeampol, Thaneerat Siripachana, Pornthep, Anussornnitisarn, 19-21 2013, zadar, Croatia, international conference
[10]. YuhanJia, Jianping Wu, and Ming Xu, Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method, Journal of Advanced Transportation, 2017.
[11]. Felix Kunde Alexander Hartenstein Stephan Pieper Petra Sauer, Traffic prediction using a Deep Learning paradigm, CEUR-WS.org, 2017
[12]. https://www.kaggle.com/fedesoriano/traffic-prediction-dataset
[13]. https://www.hindawi.com/journals/jat/2021/8878011/
[14]. https://machinelearningmastery.com/how-to-connect-model-input-data-withpredictions-for-machine-learning/

[15].https://www.shanelynn.ie/pandas-iloc-loc-select-rows-and-columns-dataframe/
[16].https://matplotlib.org/2.0.2/api/pyplot_api.html
[17].Azzouni A, Pujolle G. A long short-term memory recurrent neural network framework for network traffic matrix prediction.Comput Sci. 2017;3(6):18-27
[18].Ioannis LoumiotisRoad Traffic Prediction Using Artificial Neural Networks 2018
[19].https://www.catalyzex.com/s/Traffic%20Prediction
[20].https://www.geeksforgeeks.org/formatting-dates-in-python/
[21].https://www.scitepress.org/Papers/2016/58957/pdf/index.html
[22].Malik S., Mire A., Tyagi A.K., Arora V. (2020) A Novel Feature Extractor Based on the Modified Approach of Histogram of Oriented Gradient. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2020. ICCSA 2020. Lecture
[23].Notes in Computer Science, vol 12254. Springer, Cham. https://doi.org/10.1007/978-3030-58817-5_54
[24].B. Gudeti, S. Mishra, S. Malik, T. F. Fernandez, A. K. Tyagi and S. Kumari, "A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms," 2020 4th International Conference on Electronics, Communication and Aerospace
[25].Technology (ICECA), Coimbatore, 2020, pp. 1630-1635, doi:
[26].10.1109/ICECA49313.2020.9297392.
[27].Gillala Rekha, V. Krishna Reddy, and Amit Kumar Tyagi, “KDOS - Kernel Density based Over Sampling - A Solution to Skewed Class Distribution”, 2020, Journal of Information Assurance and Security (JIAS), Vol. 15 Issue 2, p44-52. 9p.
[28].Ambildhuke G.M., Rekha G., Tyagi A.K. (2021) Performance Analysis of Undersampling Approaches for Solving Customer Churn Prediction. In: Goyal D.,
[29].Gupta A.K., Piuri V., Ganzha M., Paprzycki M. (eds) Proceedings of the Second International Conference on Information Management and Machine Intelligence. Lecture Notes in Networks and Systems, vol 166. Springer, Singapore. https://doi.org/10.1007/978-981-15-9689-6_37
[30].Nair M.M., Kumari S., Tyagi A.K., Sravanthi K. (2021) Deep Learning for Medical Image Recognition: Open Issues and a Way to Forward. In: Goyal D., Gupta A.K., Piuri V., Ganzha M., Paprzycki M. (eds) Proceedings of the Second International Conference on Information Management and Machine Intelligence. Lecture Notes in
[31].Networks and Systems, vol 166. Springer, Singapore. https://doi.org/10.1007/978-98115-9689-6_38
[32].Amit Kumar Tyagi, Meghna Mannoj Nair, Deep Learning for Clinical and Health Informatics, in the book "Computational Analysis and Deep Learning for Medical Care:
[33].Principles, Methods, and Applications", 28 July 2021, DOI:
[34].https://doi.org/10.1002/9781119785750.ch5
[35].L. Kanuru, A. K. Tyagi, A. S. U, T. F. Fernandez, N. Sreenath and S. Mishra, "Prediction of Pesticides and Fertilizers using Machine Learning and Internet of Things," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-6, doi: 10.1109/ICCCI50826.2021.9402536.
[36].Gillala Rekha; V. Krishna Reddy; Amit Kumar Tyagi, An Earth mover's distance-based undersampling approach for handling class-imbalanced data, International Journal of Intelligent Information and Database Systems, 2020 Vol.13 No.2/3/4.
[37].Akshara Pramod, Harsh Sankar Naicker, Amit Kumar Tyagi, “Machine Learning and Deep Learning: Open Issues and Future Research Directions for Next Ten Years”, Book: Computational Analysis and Understanding of Deep Learning for Medical Care: Principles, Methods, and Applications, 2020, Wiley Scrivener, 2020.
[38].Tyagi, Amit Kumar and G, Rekha, Machine Learning with Big Data (March 20, 2019). Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur - India, February 26-28, 2019.
[39].Amit Kumar Tyagi, Poonam Chahal, “Artificial Intelligence and Machine Learning Algorithms”, Book: Challenges and Applications for Implementing Machine Learning in Computer Vision, IGI Global, 2020.DOI: 10.4018/978-1-7998-0182-5.ch008