Industrial Internship Report on

”Traffic Prediction using Machine Learning”

Prepared by

Boddula Sai Sreshta

Executive Summary

This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time.

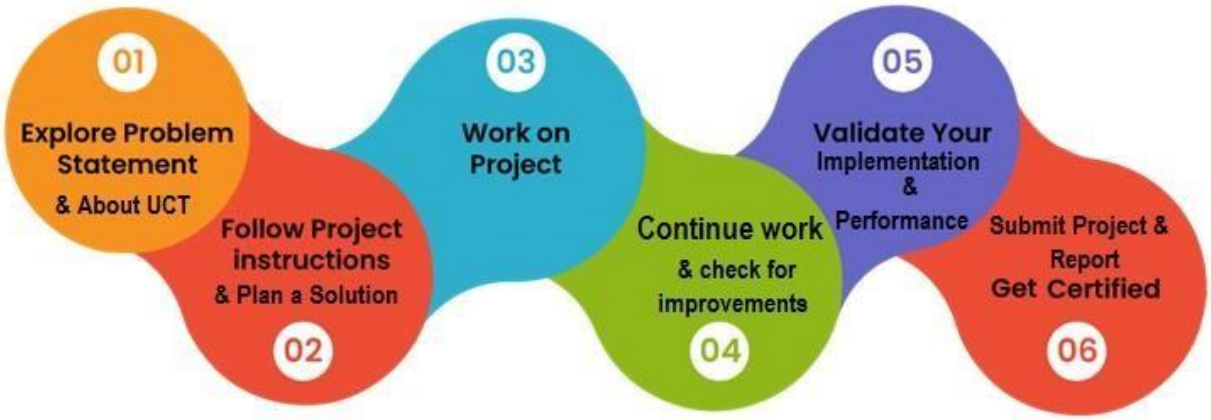
My project was Traffic Prediction using Machine Learning this prediction will be helpful for the people who are in need to check the immediate traffic state. The traffic data is predicated on a basis of 1 hour time gap. Live statistics of the traffic is analyzed from this prediction. So this will be 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. Regression model in order to predict the traffic using machine learning.

This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENT

1 Preface

Program was planned by following these steps:



My Learnings : Acquired Knowledge in DataScience and Machine Learning such as Algorithms, Types of Machine Learnings, Text mining analytics,Learning model,Prediction,Data Visualization,Probability theory&Mathematics Statistics,Linear Hypothesis,Regression Analysis and Had a Great experience in the internship.

Thankyou to team of Upskill Campus& Uniconverge Technologies.

Message to Juniors: Internship is not just about gaining experience; it's also about discovering your passions, building connections, and laying the groundwork for your future career. So embrace every opportunity, challenge yourself, and make the most of this valuable experience.

Upskill Academy is the best platform to upgrade your skills.

Best of luck on your internship journey!

2 IntroductionAbout UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

i. UCT IoT Platform (Insight)

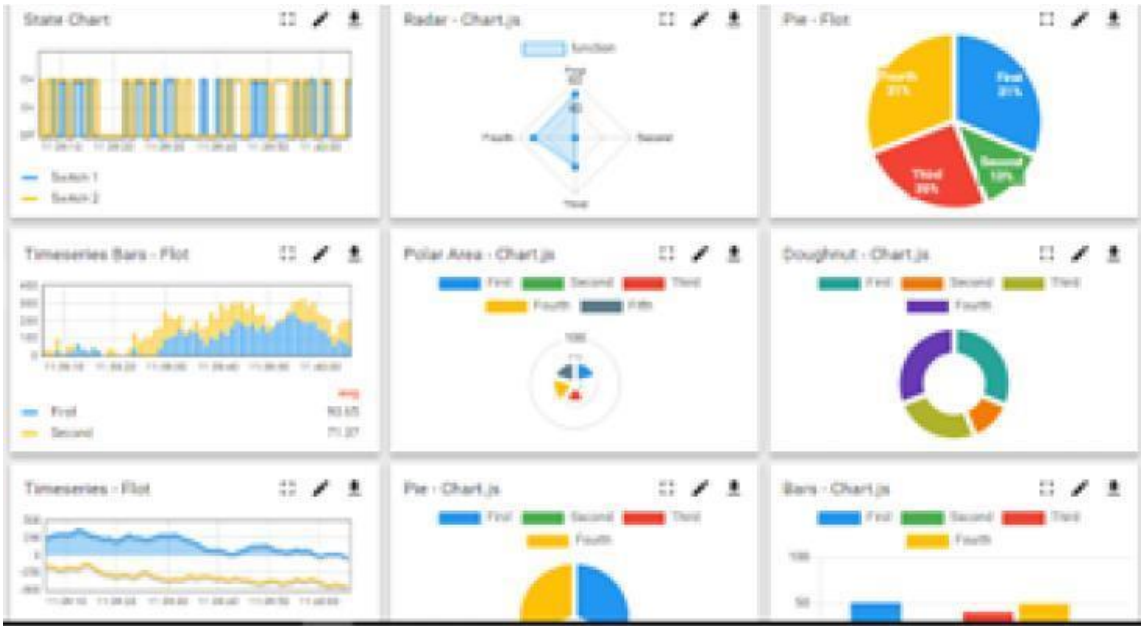
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

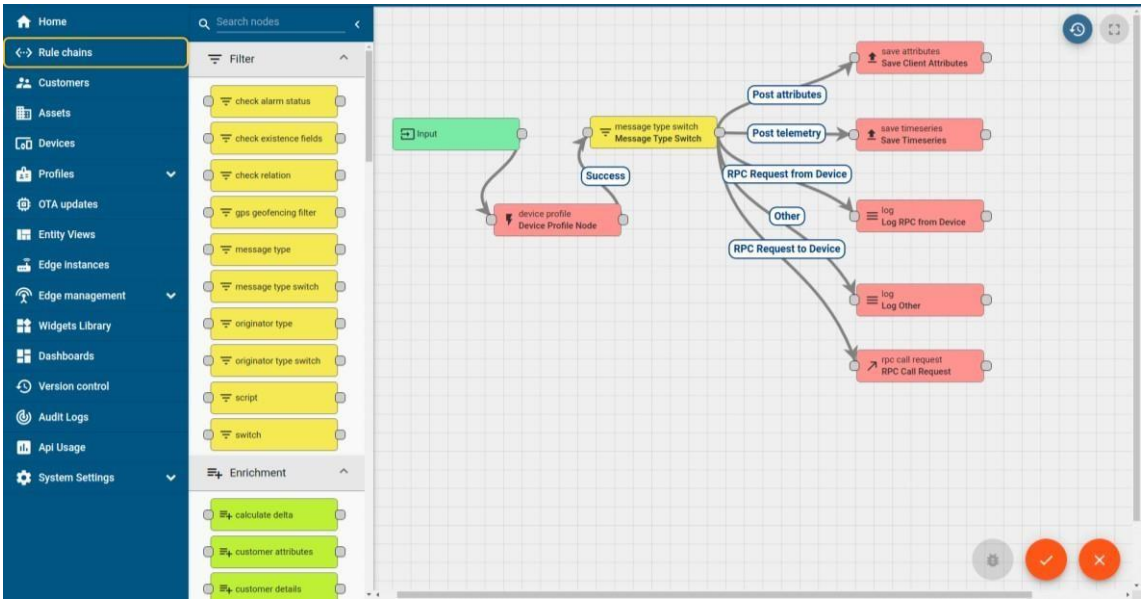
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP) • Rule Engine


State Chart
Radar - Chart.js
Pie - Flot
Timeseries Bars - Flot
Polar Area - Chart.js
Doughnut - Chart.js
Timeseries - Flot
Pie - Chart.js
Bars - Chart.js


Home
Rule chains
Customers
Assets
Devices
Profiles
OTA updates
Entity Views
Edge instances
Edge management
Widgets Library
Dashboards
Version control
Audit Logs
Api Usage
System Settings
Search nodes
Filter
check alarm status
check existence fields
check relation
gps geofencing filter
message type
message type switch
originator type
originator type switch
script
switch
Enrichment
calculate delta
customer attributes
customer details
Input
message type switch
Message Type Switch
device profile
Device Profile Node
Success
save attributes
Save Client Attributes
Post attributes
Post telemetry
save timeseries
Save Timeseries
RPC Request from Device
log
Log RPC from Device
Other
RPC Request to Device
log
Log Other
rpc call request
RPC Call Request

ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

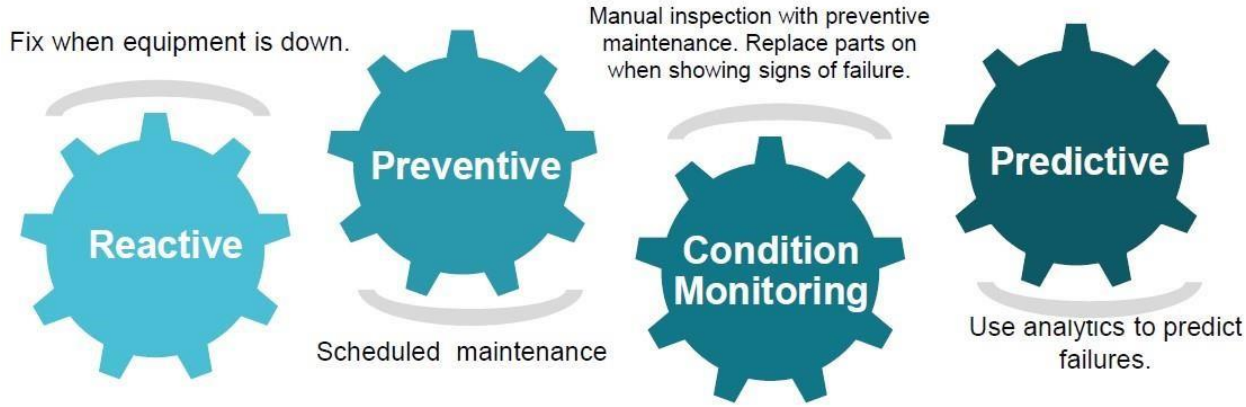
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

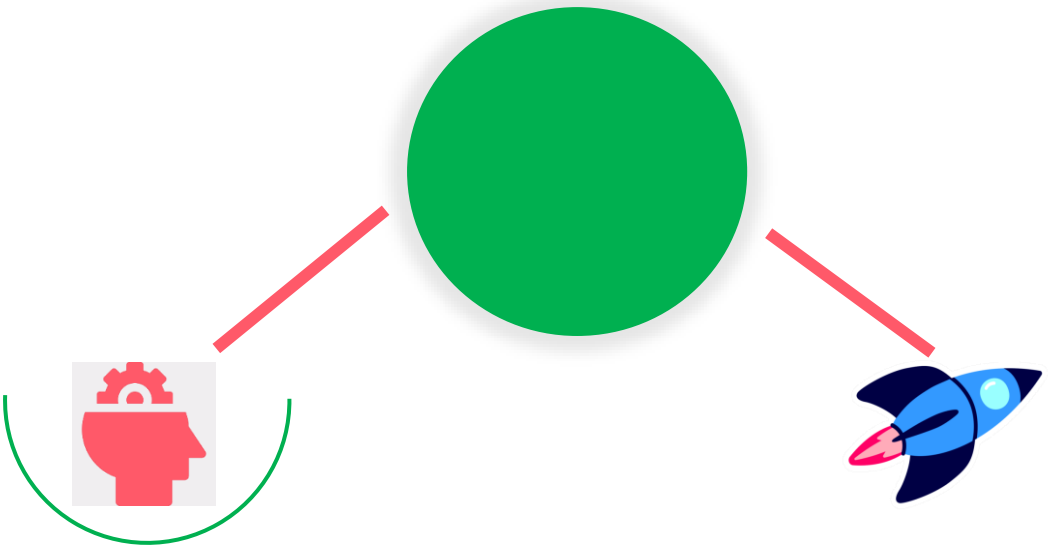
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.

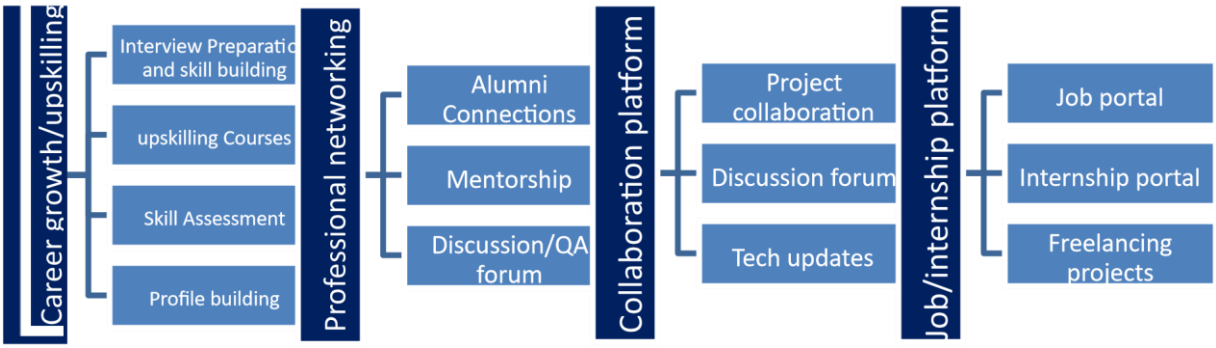


2.1 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.





2.2 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

Seeing need of upskilling in self paced manner along -with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/

2.3 Objectives of this Internship program

The objective for this internship program was to ☛ get

practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

3. Problem Statement:

To overcome the problem of traffic congestion, the traffic prediction using machine learning which contains regression model and libraries like pandas, os, numpy, matplotlib.pyplot are used to predict the traffic. This has to be implemented so that the traffic congestion is controlled and can be accessed easily. Users can collect the traffic

information of the traffic flow and can also check the congestion flow from the start of the day till the end of the day with the time span of one hour data. In this way, Users can know the weather conditions of the roads that they would probably opt to take. This also tells the accuracy of the traffic by comparing their mean square errors of the past year's data and the recent year's data. Users can also know how many vehicles are traveling on average by the traffic prediction.

4. Purpose of statement

Many reports of the traffic data are of actual time but it is not favorable and accessible to many users as we need to have prior decision in which route we need to travel. For example, During working days, we need to have daily traffic information or at times we need hourly traffic information but then the traffic congestion occurs; for solving this issue the user need to have actual time traffic prediction. Many factors are responsible for the traffic congestion. This can be predicted by taking two datasets; one with the past year and one with the recent year's data set. If traffic is so heavy then the traffic can be predicted by referring the same time in the past year's data set and analyzing how congested the traffic would be. With the increasing cost of the fuel, the traffic congestion changes drastically. The goal of this prediction is to provide real-time gridlock and snarl up information. The traffic on the city becomes complex and are out of control these days, so such kind of systems are not sufficient for prediction. Therefore, research on traffic flow prediction plays a major role in ITS.

5 .Proposed Design Model

Traffic congestion is raising a lot these days. Factors like expanding urban 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.

5.1 Regression model

Regressor model analysis could even be a mathematical technique for resolving the connection in the middle of one dependent (criterion) variable and one or more independent (predictor) variables.

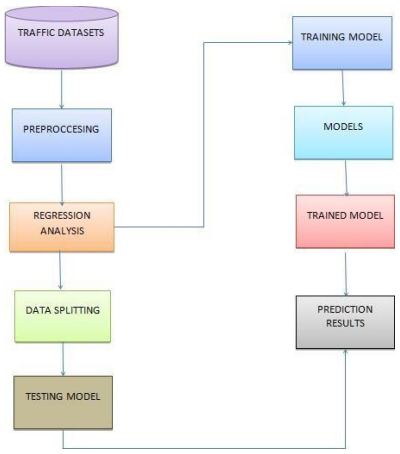


Figure 1. Regression model of traffic prediction

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 method .Figure 1 shows the Regression model for the Traffic Prediction.

JupyterLab is a browser based communal development. JupyterLab is a limber and which can construct and exhibit the user interface to support a far flung of metadata in machine learning.Python3 is the status quo environment where the code is implemented in Jupyter notebook. This can be accesses/installed using command prompt. This is done in order to get the access from the local drive. So the Jupyter notebook is installed through command prompt and then a local host is created. The file is accessed though this host and the prediction are done using various libraries and models in the python environment.

6.Performance Test

6.1Software Implementation

Simulation

The command prompt is the local host in this paper to initialize the jupyter notebook.

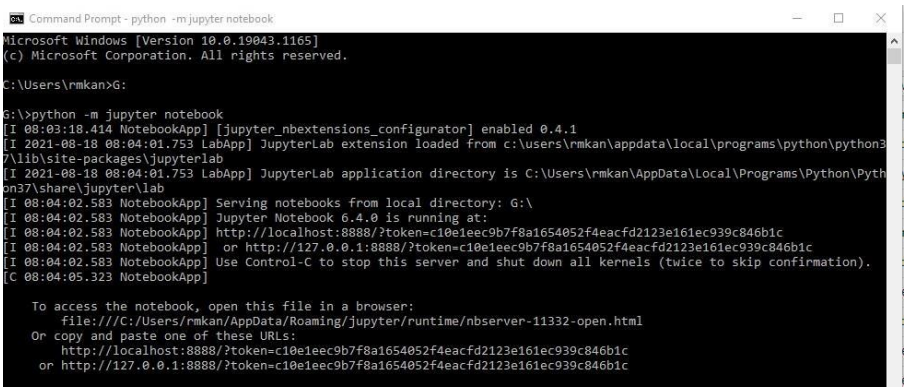


Figure 2. The figure signifies the initializing the jupyter notebook through command prompt

The local host contains the nbextenisons which we modify to our convenience

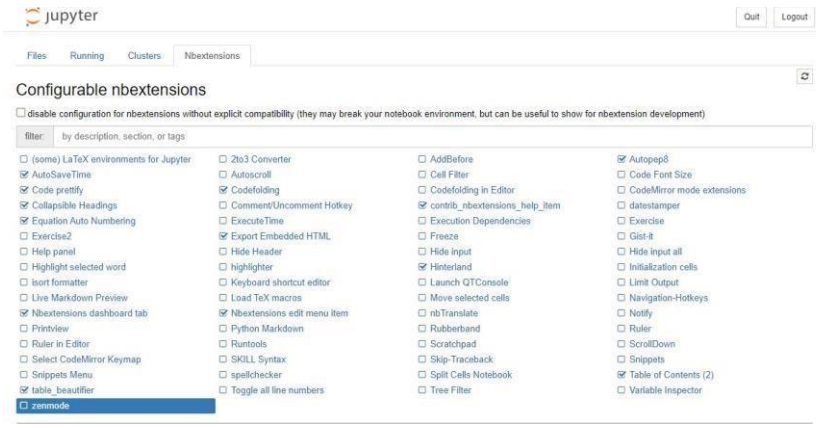


Figure 3 The figure denotes the necessary nbextenisons that is needed for the prediction to take place

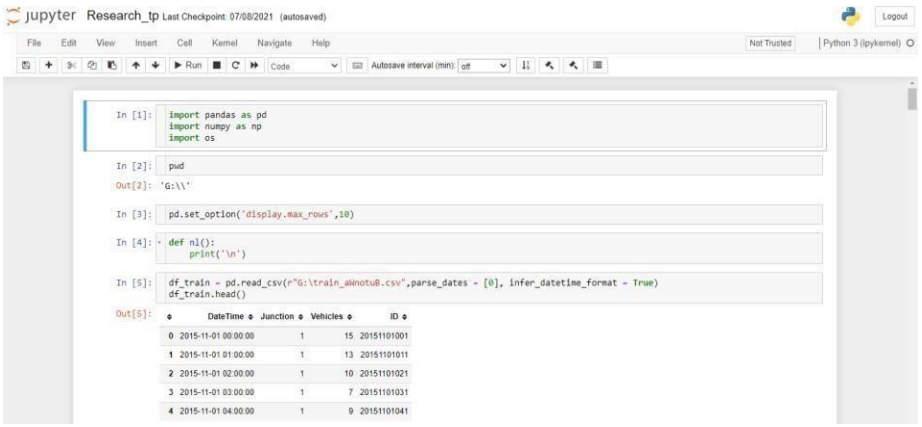


Figure 4:Implies the page of the jupyter notebook.

```
In [6]: df_test = pd.read_csv(r"G:\test_BdBKkAj.csv",parse_dates = [0], infer_datetime_format = True)
        df_test.head()
```

Out[6]:

	DateTime	Junction	ID
0	2017-07-01 00:00:00	1	20170701001
1	2017-07-01 01:00:00	1	20170701011
2	2017-07-01 02:00:00	1	20170701021
3	2017-07-01 03:00:00	1	20170701031
4	2017-07-01 04:00:00	1	20170701041

Figure 5. The figure above denotes the result after running the particular command in the prediction process will be displayed under out command

Code submission(GitHub link)-
https://github.com/sreshta121/Upskill/blob/main/trafficpredictionusingmachinelearning.pythonML

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]

	DateTime	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

DateTime	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
2017-07-01 04:00:00	1	20170701041

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
2017-06-30 23:00:00	78.0	27.0	39.0	12.0

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

COMMANDS ARE USES:

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

7.Results from the simulation

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

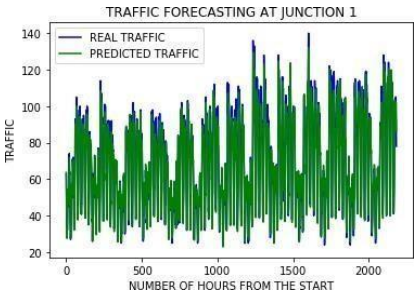


Figure 6. The figure signifies the Traffic prediction of Junction 1 from the datasets

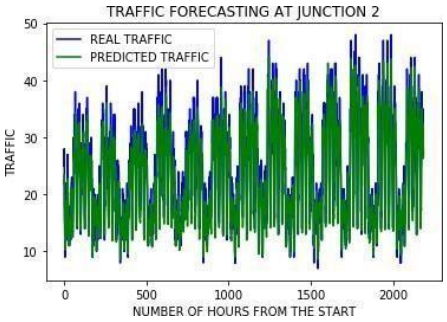


Figure 7. The figure signifies the Traffic prediction of Junction 2 from the datasets

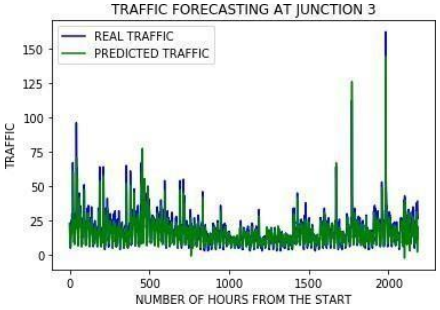


Figure 8. The figure signifies the Traffic prediction of Junction 3 from the datasets

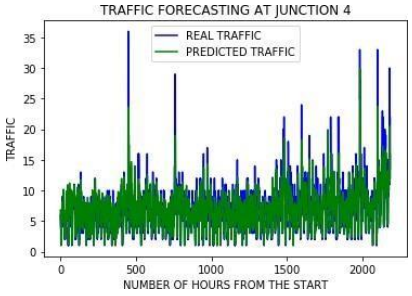


Figure 9. The figure signifies the Traffic prediction of Junction 4 from the datasets

8. 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 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 car pooling, 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.

9.Future Work Scope

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 accidentprone 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[18].

10.References

- 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.
- Big data-driven machine learning-enabled traffic flow prediction Fanhui Kong1 Jian T https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2845248/

3.https://jupyter.org/

4.Bao G, Zeng Z, Shen Y. Region stability analysis and tracking control of memristive recurrent neural network. Neural Netw. 2017;5(1):74-89.