Traffic Volume Estimation Using IBM Watson Machine Learning

1.INTRODUCTION

1.1Overview

This project aims at building an Application that automatically estimates traffic volume by taking the input values.Growth in the number of vehicles and degree of urbanization means that the annual cost of traffic jams is increasing in cities. This leads to a decrease in the quality of life among citizens through a considerable waste of time and excessive fuel consumption and air pollution in congested areas.

Traffic congestion has been one of the major issues that most metropolises are facing despite measures being taken to mitigate and reduce it. The safe and time-efficient movement of the people and goods is dependent on Traffic flow, which is directly connected to the traffic characteristics. Early analysis of congestion events and prediction of traffic volumes is a crucial step to identify traffic bottlenecks, which can be utilized to assist traffic management centres. We will be using Regression algorithms such as Linear Regression, Decision tree, Random forest, and xgboost to predict the count of traffic volume. We will train and test the data with these algorithms. From this best model is selected and saved in .pkl (Pickle) format. Once the model is saved, we integrate it with flask application and also deploy the model in IBM.

1.2 Purpose

The traffic volume studies are conducted to obtain accurate information about the number and movement of vehicles within or through an area or at a selected points within the area. The accuracy level of the ML models used in predicting Traffic volume based on historical data has been one of the most critical concerns in Transportation studies. An accurate ML model could give early alerts of severe traffic to help prevent issues related to traffic problems. Hence, there is needs to develop ML algorithms capable in predicting Traffic volume with acceptable level of precision and in reducing the error in the dataset of the projected Traffic volume from model with the expected observable Traffic volume.An accurate prediction model can help drivers optimize the routes allocate resources reasonably an reduce urban traffic congestions.

2.LITERATURE SURVEY

2.1 Exsisting problem

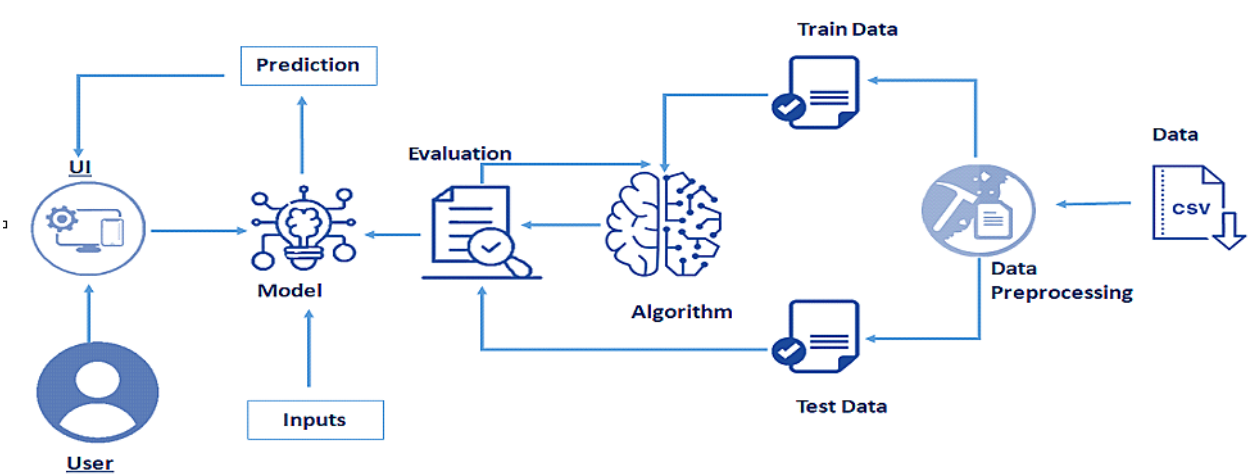
Traditional approaches for traffic volume estimation and prediction heavily rely on data from various road-based sensors that is, loop dectectors or surveillance cameras, thus mainly applicable to major road sections or limited-scale road networks.In many of these studies ,traffic volume measures are directly monitered by sensors, and the volume estimation or prediction is achieved using filtering based algorithms.

2.2 Proposed System

This project aims at building an Application using ibm Watson Machine learning that automatically estimates the traffic volume by taking the input values.The studies are conducted to obtain accurate information about the number and movement of vehicles within or through an area or at a selected points within the area. We will be using regression algorithms such as Linear Regression, Decision tree, Random forest, and xgboost to predict the count of traffic volume. We will train and test the data with these algorithms. From this best model is selected and saved in .pkl (Pickle) format. Once the model is saved, we integrate it with flask application and also deploy the model in IBM.

3.THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware/Software designing

IBM Watson Studio-IBM Watson Studio helps data scientists and analysts prepare data and build models at scale accross any cloud.

IBM Watson Machine Learning-IBM Watson Machin Learning helps data scientists and devolopers accelerate AI and machine-Learning deployment.

IBM Cloud Object Storage-IBM Cloud Object Storage makes it possible to store practically limitless amount of data,simply and cost effectively.

4.EXPERIMENTAL INVESTIGATION

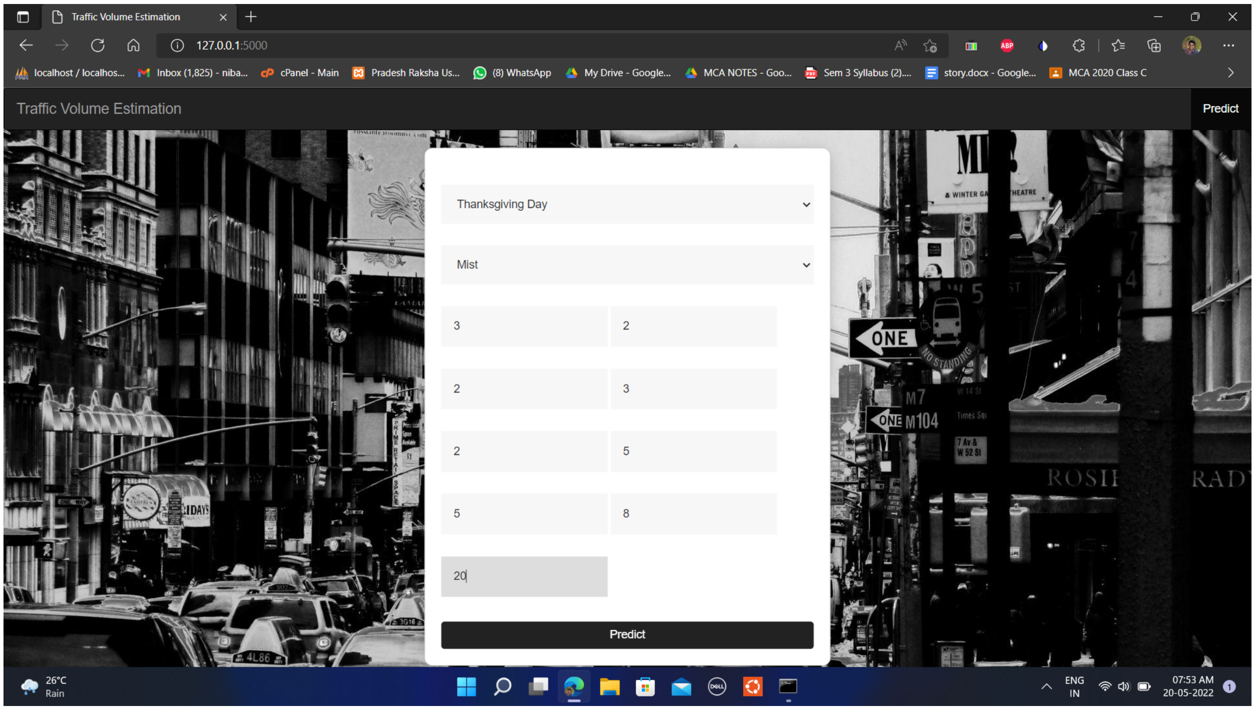
Here we are going to build a machine learning model that predicts Insuranceprediction based on the following parameters,

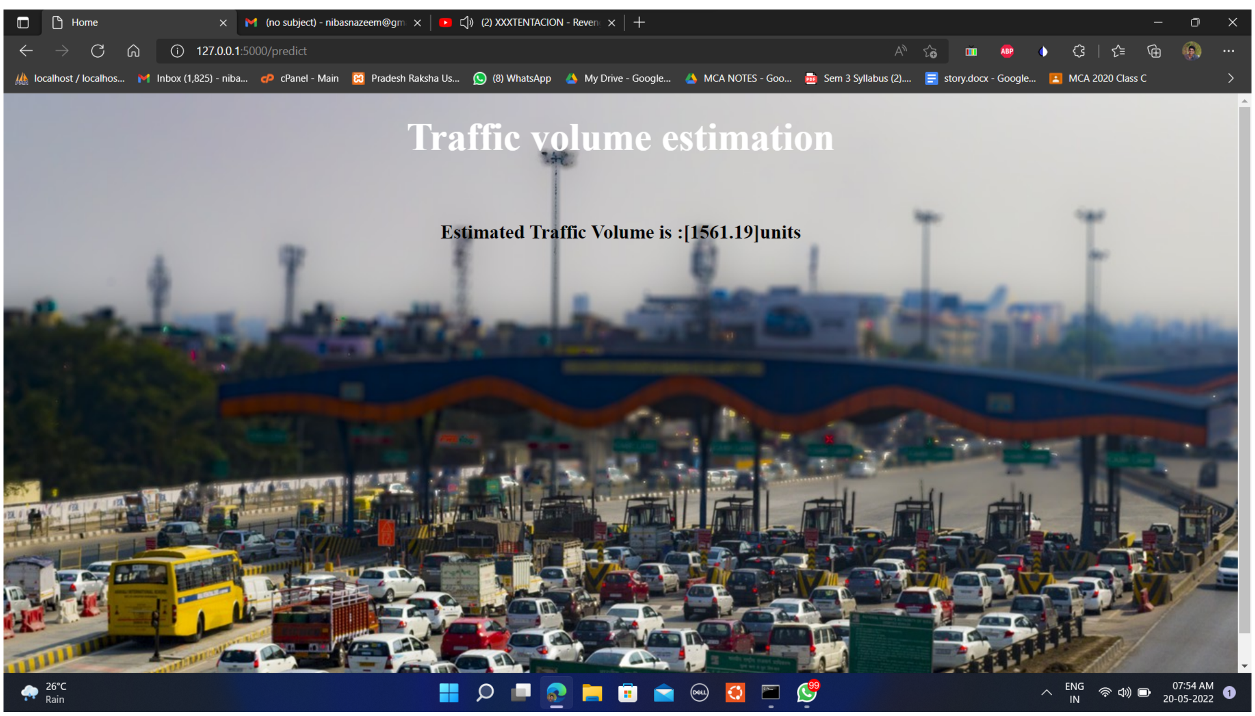
- day
- temperature
- rain
- snow
- weather

Here there are 5 parameters which affect the volume of traffic.They are the type of holiday ,temperature,rain,snow and weather. All these values are the factors for predicts the traffic volume.

5.RESULT

When we input the value in each input fields as our wish for an instance and
we got output as,


Traffic Volume Estimation
127.0.0.1:5000
Traffic Volume Estimation
Predict
Thanksgiving Day
Mist
3
2
2
3
2
5
5
8
20
Predict


Home
127.0.0.1:5000/predict
Traffic volume estimation
Estimated Traffic Volume is :[1561.19]units

6.ADVANTAGES AND DISADVANTAGES

6.1 Advantages

- System allows automate traffic volume prediction
- Allows for faster Service
- Time Saving
- The interface is user friendly

6.2 Disadvantages

The risks foreseen in the project is that the machine need to be trained with large datasets in order to obtain the accurate prediction of the traffic volume, if the dataset is reduced the accuracy is also reduced. Also, the system configuration is a major factor in the project and therefore 100% accuracy is not assured.

7.APPLICATIONS

The traffic volume prediction is usefull for many applications and its very important for devolopment of better infrastructure to the cities and finally to the country.

- It increase the efficiency and life of roads
- Reduce traffic volume at a particular sector
- Provide beter means for devolopment of infrastructures
- Provide better means to utilize other roads in case of special events in the city

8.CONCLUSIONS

Traffic volume prediction is getting more attention from last few decades. With the devolopment of infrastructure, every country is facing traffic congestion problem. Therefore, forecasting the congestion can allow authorities to make plan and take necessary actions to avoid it. The development of artificial intelligence and the availability of big data have led to apply different models in this field. Although for devolping models different factors that affect traffic congestion, e.g., weather,day,temperature,rain snow etc are considered and predicting accurate measurement.However, a wide range of machine learning algorithms are yet to be applied. Therefore, the vast opportunity of research in the field of traffic congestion prediction still prevails.

9.FUTURE SCOPE

The successfull machine learning system that manages the traffic volume is its ability to devolop efficient reasoning and intutively read and understand the trends. so in the future we can add more parameters to predict and analyze the traffic volume estimation.

10.BIBLIOGRAPHY

- https://www.osti.gov/biblio/1669474-ubiquitous-traffic-volume-estimation-through-machine-learning-procedure
- https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa17036.pdf
- https://www.researchgate.net/publication/309445662_Citywide_Traffic_Volume_Estimation_Using_Trajectory_Data

11.APPENDIX

Source code

```
import pandas as pd
import numpy as np
import seaborn as sns
import sklearn as sk
from sklearn import linear_model
from sklearn import tree
from sklearn import ensemble
from sklearn import svm
import xgboost
import matplotlib
Data=pd.read_csv(r'C:\traffic volume.csv')
Data.head()
Data.describe()
Data.shape
Data.info()
#dealing with missing values
Data.isnull().sum()
Data.dtypes
from collections import Counter
print(Counter(Data['rain']))
print(Counter(Data['snow']))
Data['temp'].fillna(Data['temp'].mean(),inplace=True)
Data['rain'].fillna(Data['rain'].mean(),inplace=True)
Data['snow'].fillna(Data['snow'].mean(),inplace=True)
print(Counter(Data['weather']))
```

```
Data['weather'].fillna('Clouds',inplace=True)
Data.isnull().sum()
#Datavisualization
Data.corr()
sns.pairplot(Data)
Data.boxplot()
#ENCODING THE DATA
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
data['weather'] = le.fit_transform(data['weather'])
Data['holiday'] = le.fit_transform(Data['holiday'])
Data.head()
# spliiting the date column into year,month,day
data[["day", "month", "year"]] = data["date"].str.split("-", expand = True)
data[["hours", "minutes", "seconds"]] = data["Time"].str.split(":", expand = True)
data.drop(columns=['date','Time'],axis=1,inplace=True)
data.head()
#splitting the dataset into dependent and independent variable
y = Data['traffic_volume']
x = Data.drop(columns=['traffic_volume'],axis=1)
names = x.columns
from sklearn.preprocessing import scale
x = scale(x)
x  = pd.DataFrame(x,columns=names)
x.head()
 #splitting the data into train and set
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
#MODEL BUILDING
from sklearn import linear_model
from sklearn import tree
from sklearn import ensemble
from sklearn import svm
import xgboost
lin_reg = linear_model.LinearRegression()
Dtree = tree.DecisionTreeRegressor()
Rand = ensemble.RandomForestRegressor()
svr = svm.SVR()
XGB = xgboost.XGBRegressor()
lin_reg.fit(x_train,y_train)
Dtree.fit(x_train,y_train)
```

```
Rand.fit(x_train,y_train)
svr.fit(x_train,y_train)
XGB.fit(x_train,y_train)
p1 = lin_reg.predict(x_train)
p2 = Dtree.predict(x_train)
p3 = Rand.predict(x_train)
p4 = svr.predict(x_train)
p5 = XGB.predict(x_train)
#model evaluation
from sklearn import metrics
print(metrics.r2_score(p1,y_train))
print(metrics.r2_score(p2,y_train))
print(metrics.r2_score(p3,y_train))
print(metrics.r2_score(p4,y_train))
print(metrics.r2_score(p5,y_train))
#with testing data finding the rscore
p1 = lin_reg.predict(x_test)
p2 = Dtree.predict(x_test)
p3 = Rand.predict(x_test)
p4 = svr.predict(x_test)
p5=XGB.predict(x_test)
print(metrics.r2_score(p1,y_test))
print(metrics.r2_score(p2,y_test))
print(metrics.r2_score(p3,y_test))
print(metrics.r2_score(p4,y_test))
print(metrics.r2_score(p5,y_test))
MSE = metrics.mean_squared_error(p3,y_test)
np.sqrt(MSE)
import pickle
pickle.dump(Rand,open("model.pkl",'wb'))
pickle.dump(le,open("encoder.pkl",'wb'))
```