PROJECT REPORT

Traffic Telligence: Advance Traffic volume Estimation With Mchine Learning

1.INTRODUCTION

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 areasTraffic 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. 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

1.1PROJECT OVERVIEW

Many state and local agencies are currently facing challenges concerning the collection and estimation of traffic volumes, particularly regarding the collection of annual average daily traffic (AADT) on low-volume roads. To overcome these challenges, there is a need to develop new affordable methods to collect data and estimate traffic volume on low-volume roadways. In this study, the research team developed an innovative interpretable machine learning framework and applied it to low-volume roads in Vermont to estimate traffic volumes. This study used several databases (e.g., U.S. Census, the American community survey) to prepare the final dataset for the model development. The findings show that population density and work area characteristic (WAC) density are the best predictors in estimating AADT. The model outcomes show that the machine learning models yield better estimates than the conventional parametric statistical methods. By improving the accuracy of AADT estimations, this study contributed to traffic monitoring and safety improvement, and it can help reduce costs of data collection. This study developed the top five decision rules for three types of low-volume roadways. Stakeholders can use the findings of this study to meet the new requirements pertaining to availability of AADT estimates for low-volume roads. Additionally, the best fit estimates and the developed rules from the current study could enhance the predictive power of the SPF development for the low-volume roadways in Vermont and therefore improve the decision-making process.

1.2PURPOSE

Traffic volume studies are conducted to determine the number, movements, and classifications of roadway vehicles at a given location. These data helps to identify critical flow time periods, determining the influence of large vehicles or pedestrians on vehicular traffic flow.

2.IDEATION & PROPOSED SOLUTION

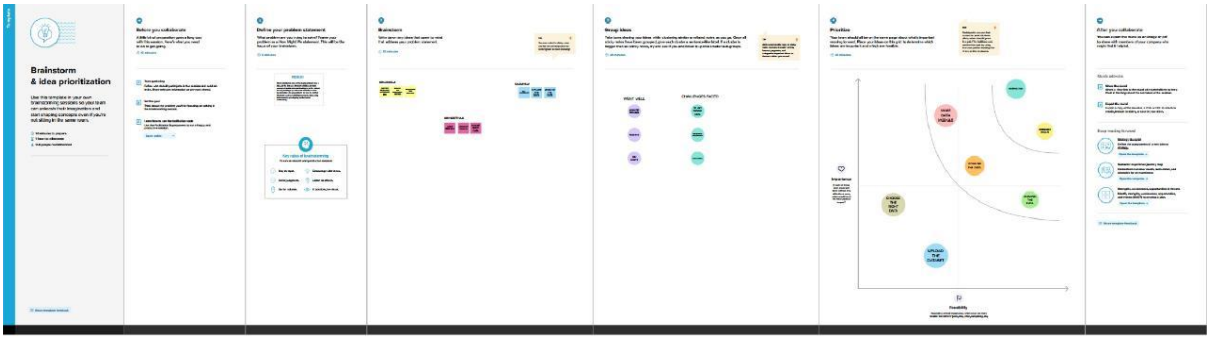
2.1 Problem Statement Definition

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Working employees	Know the best ways to reach the place in fast ways	There is proper information about it	It takes more time to properly find them	Annoyed
PS-2	Doctor	Want to know the fastest way to reach the place.	It is difficult to sort out	There is no proper data.	Uncomfortable
PS-3	Travel agency	Want to know the best routes in urban, rural areas	Finding each one separately is harder	Lack of data	Frustrated
PS-4	Traffic policeman	Reduce the traffic	It turns out that there is more of digital way of reducing traffic	It Predict traffic well	Happier

2.2Empathy Map Canvas

Pain	Gain
Undetected movements Time predicition	Saves our time Comfortable to use

2.3 Ideation & Brainstorming



2.4 Proposed Solution

1.	Problem Statement (Problem to be solved)	Reducing the traffic, congestion and improving safety on roadways
2.	Idea / Solution description	Traffic volume estimation is a crucial component of transportation planning and management. Accurate and reliable estimation of traffic volume is essential for optimizing traffic flow, reducing congestion, and improving safety on roadways. Machine learning (ML) techniques have been widely used in recent years for traffic volume estimation due to their ability to handle complex and large-scale data sets
3.	Novelty / Uniqueness	Accurate and reliable estimation of traffic volume is obtained
4.	Social Impact / Customer Satisfaction	Improved Safety: Accurate traffic volume estimation can help identify high-risk areas Reduced Congestion: Traffic volume estimation can help optimize traffic flow and reduce congestion
5.	Business Model (Revenue Model)	Logistics: Traffic volume estimation can help logistics companies optimize their delivery routes and schedules, reducing transportation costs and improving delivery times. Businesses require real-time monitoring of traffic data to make informed decisions. This includes monitoring traffic flow, congestion, accidents, weather conditions, and other factors that may impact traffic
6.	Scalability of the Solution	Provides accurate results and detailed information about the traffic in the particular places.

3.REQUIREMENTS ANALYSIS

3.1Functional requirements

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
1	**User Registration**	Registration through Gmail
2	**User Confirmation**	➢ Confirmation via Email ➢ Confirmation via OTP
3	**User Interface**	Create your profile and enter the essential details that is required
4	**User Input**	➢ Upload the dataset ➢ Upload the data, day,time,location
5	**Data Processing**	➢ Evaluate the model using test dataset ➢ Train the dataset by Random forest algorithm
6	**Report Generation**	The final report with the Traffic Volume Estimation for each person will be displayed

3.2 Non functional requirements

FR No.	Non-Functional Requirement	Description
1	**Usability**	➢ Used to identify the easy routes ➢ Classify the rotes with less traffic using Machine learning ➢ Can be used for all type of people
2	**Security**	➢ The received data should be kept and processed confidentially ➢ The data can be accessed only by authorized people
3	**Reliability**	By the usage of efficient algorithm, it must classify the less traffic routes without any mistakes.
4	**Performance**	➢ Data of different people can be processed at the same time ➢ Greater accuracy in result ➢ Saves time
5	**Availability**	➢ It is available to anyone with access to internet.
6	**Scalability**	➢ Number of dataset processing won't affect the performance of the system. ➢ Thus, it can process the data of large number of routes and produce accurate results.

4.PROJECT DESIGN

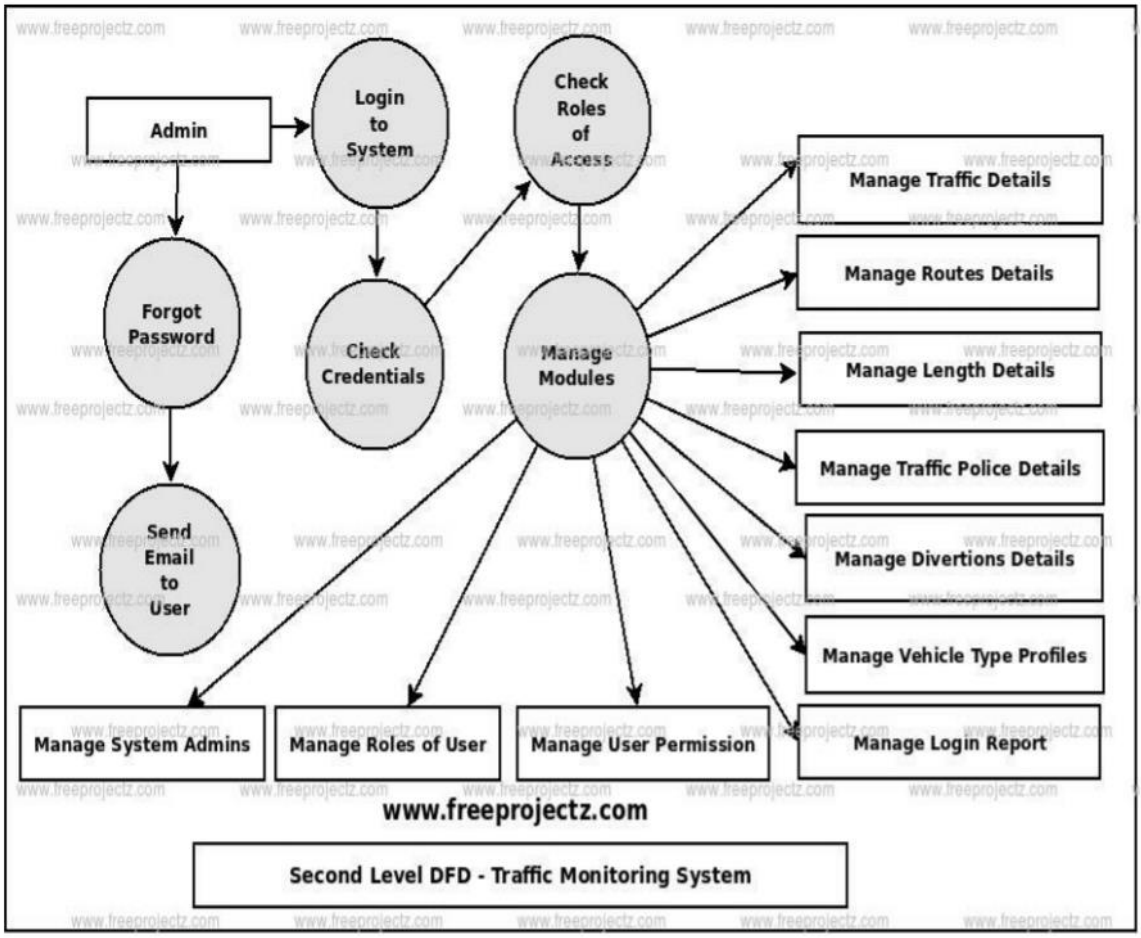
4.1 Data Flow Diagram

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right

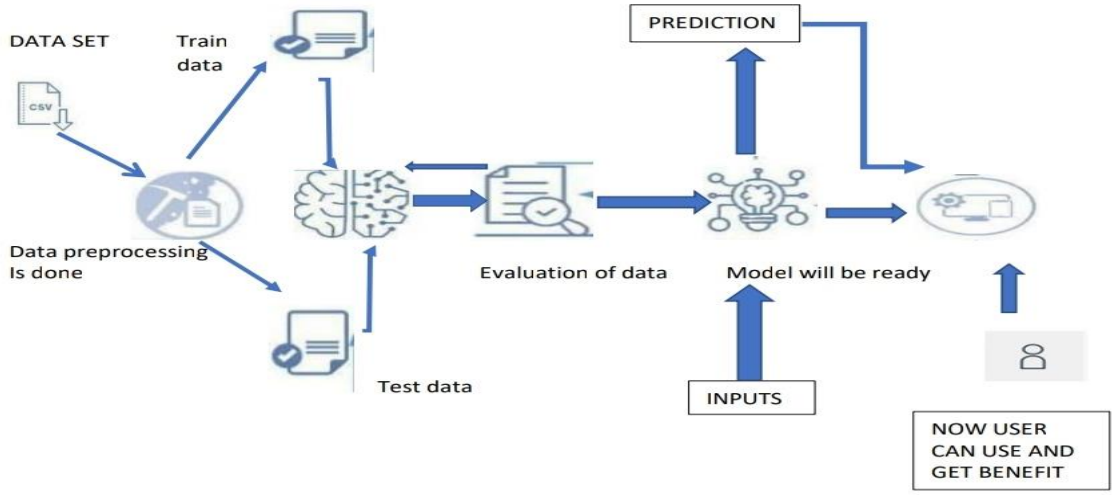
amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is

stored.

Example: DFD Level 2 (Industry Standard)



4.2 Solution & Technical Architecture



4.3 User Stories

User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Srilekha
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	kaaviya
		USN-3	As a user, I can register for the application through Gmail		Medium	Sangeetha
	Login	USN-4	As a user, I can log into the application by entering email & password		High	kaaviya
	Dashboard	USN-5	As a user can view the sample project phase template and documents	I can access dashboard	High	Sangeetha
Customer (Web user)	WebPortal	USN-6	Customer can login and visit the dataset	Receive verification code	Medium	srilekha
Customer Care Executive	OnlineVisit	USN-7	Ask if any queries arrises while performing tasks		Low	kaaviya
Administrator	Maintenance	USN-8	Administrator can maintain data frame set and database		High	srilekha

5.CODING & SOLUTIONING

- Importing The Data Set

might have your data in .csv files, .excel filesLet's load a .csv data file into pandas using read_csv() function.We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).If your dataset is in some other location, Then**Data=pd.read_csv(r"File_location/datasetname.csv")**

```
# importing the data
data = pd.read_csv(r"G:\AI&ML\ML projects\Traffic_volume\traffic volume.csv")
```

Note: r stands for "raw" and will cause backslashes in the string to be interpreted as actual backslashes rather than special characters.
If the dataset is in the same directory of your program, you can directly read it, without giving raw as r.Our Dataset weatherAus.csv contains the following ColumnsHoliday - working day or holidayTemp- temperature of the dayRain and snow – whether it is raining or snowing on that day or notWeather = describes the weather conditions of the dayDate and time = represents the exact date and time of thedayTraffic volume – output column.The output column to be predicted is Traffic volume.Based on the input variables we predict the volume of the traffic. The predicted output gives them a fair idea of the count of traffic

ANALYZE THE DATA

head() method is used to return top n (5 by default) rows of a DataFrame or series

```
data.head()
```

	holiday	temp	rain	snow	weather	date	Time	traffic_volume
0	None	288.28	0.0	0.0	Clouds	02-10-2012	09:00:00	5545
1	None	289.36	0.0	0.0	Clouds	02-10-2012	10:00:00	4516
2	None	289.58	0.0	0.0	Clouds	02-10-2012	11:00:00	4767
3	None	290.13	0.0	0.0	Clouds	02-10-2012	12:00:00	5026
4	None	291.14	0.0	0.0	Clouds	02-10-2012	13:00:00	4918

describe() method computes a summary of statistics like count, mean, standard deviation, min, max, and quartile values.

```
data.describe()
```

MODEL BUILDUING

The model building includes the following main tasks

Import the model building Libraries

Initializing the model

Training and testing the model

Evaluation of Model

Save the Model

SPILLING THE X-TRAIN AND Y-TRAIN

- Once after splitting the data into train and test, the data should be fed to an algorithm to build a model.
- There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have it may be Classification algorithms are Regression algorithms.
  1. Linear Regression
  2. Decision Tree Regressor
  3.Random Forest Regressor
  4.svm
  5.xgboost

Steps in Building the model:-

Initialize the model -

```
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()
```

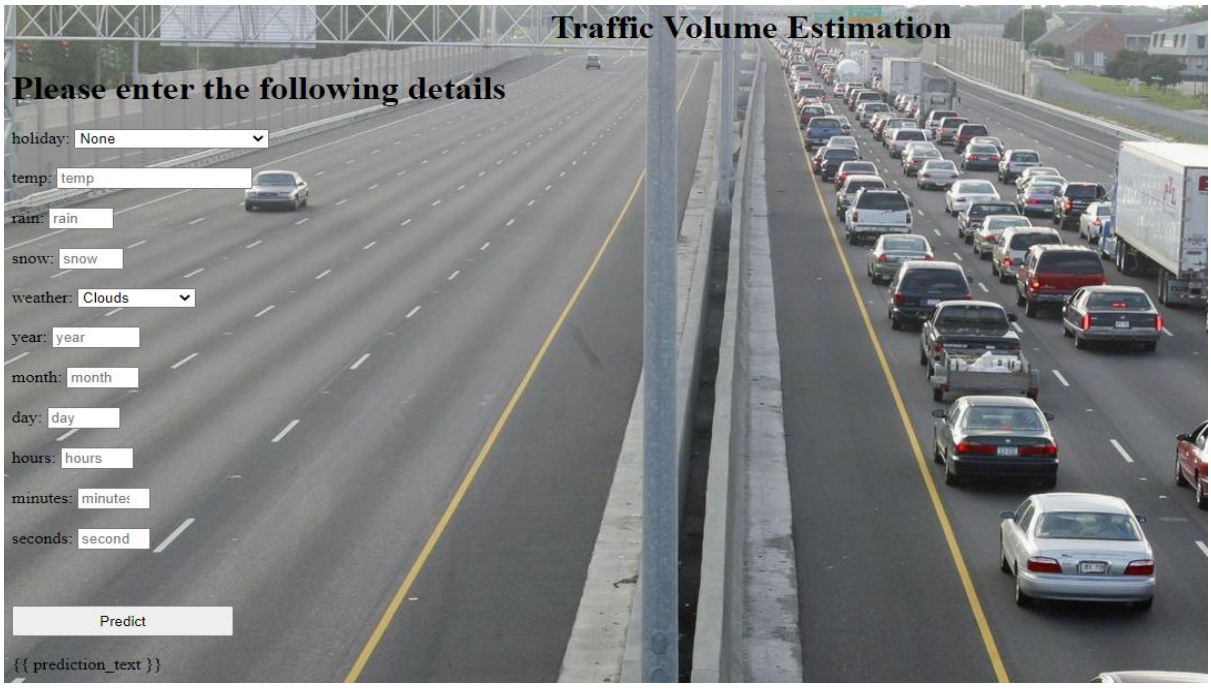
Fit the models with x_train and y_train -

```
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)
```

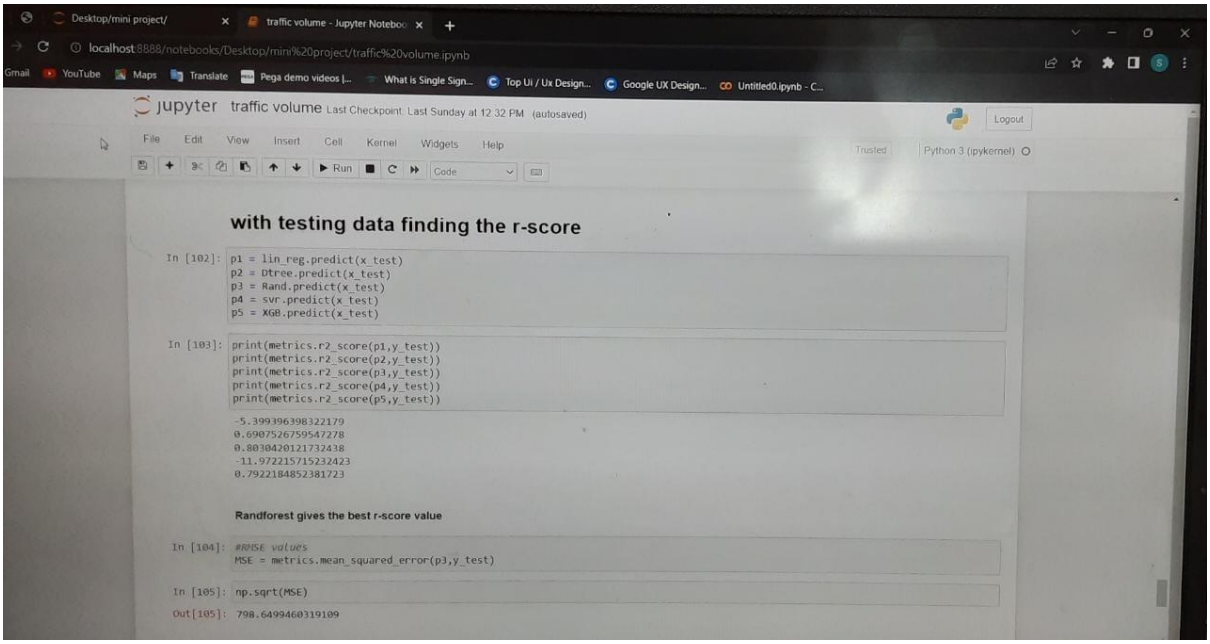
Predict the y_train values and calculate the accuracy -

```
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)
```

6.RESULT



6.1 Performance Metrics :



7.ADVANTAGES & DISADVANTAGES

Advantages: By this method traffic volume as well as vehicle classification and turning proportions can be obtained. Data can be used immediately after collection.

Disadvantages: This method is not practicable for long duration count and when flow is high.

8.CONCLUSION

Traffic volume data analysis is important for many transportation research areas, , including roadway safety improvement and design, countermeasure determination, travel model calibration and validation, pavement design, and air quality compliance.
However, AADT data are more easily available for higher functional classes, and only a small percentage of low-volume roads have accurate AADT data. Because low-volume roadways constitute a large portion of the U.S. roadway network, more studies are needed in traffic volume prediction on these roadways.
The main contribution of this study was the development of a robust interpretable machine learning framework that can be used to estimate AADT; this framework can be adopted by other researchers and practitioners. The findings of this study show that machine learning models can perform better than conventional linear regression models. Additionally, the research team found that population and work employment density are the best predictors for all three low-volume roadway classes in Vermont. In this study, the best fit machine learning model (random forest) has higher R^2 values in comparison to the statistical models. Compared with results from regression models, the best fit random forest model improved the accuracy of AADT for low-volume roadways significantly, from 0.45 to 0.77. The partial dependent plots developed for combination of variables show different clusters with the estimated AADT values. This study developed the top five decision rules for three functional classes of roadways. The best fit estimates and the developed rules from the current study could enhance the predictive power of the SPF development for the low-volume roadways in Vermont and therefore improve the decision-making process.
The current study does have a few limitations. The main limitation is the small sample size for two of the low-volume roadways (rural collector or 6R and urban local or 7U). With access to a larger sample size, the model performance can be improved. Additionally, the current framework is machine learning based, so structural equations are not available for interpretation of the results. However, PDPs are alternative tools which can transform the black box algorithm into interpretable estimations.

9.FUTURE SCOPE

The present study is focused mainly on traffic volume only. Speed-flow studies are useful to evaluate the more parameters. There is a scope on speed flow studies on urban road links for future work.

It's important to note that while these future scopes hold great potential, their implementation and widespread adoption may take time due to various factors such as technological advancements, infrastructure requirements, and policy considerations. However, as technology continues to evolve, we can expect significant improvements in traffic volume estimation, leading to more efficient transportation systems**.**