**Ideation Phase**

**Brainstorm & Idea Prioritization Template**

|  |  |
| --- | --- |
| Date | 15 June 2025 |
| Team ID | LTVIP2025TMID41256 |
| Project Name | Traffic Telligence Advanced Traffic Volume  Estimation With Machine Learning |
| Maximum Marks | 4 Marks |

TrafficTelligence is an advanced system that uses machine learning algorithms to estimate and predict traffic volume with precision. By analyzing historical traffic data, weather patterns, events, and other relevant factors, TrafficTelligence provides accurate forecasts and insights to enhance traffic management, urban planning, and commuter experiences.

"Accurate traffic volume estimation is crucial for optimizing traffic management, reducing congestion, and improving travel times. However, traditional methods often rely on manual counting, limited sensor data, and simplistic models, leading to inaccurate estimates and inefficient resource allocation.

"This is where TrafficIntelligence comes in – harnessing the power of Artificial Intelligence (AI) and Machine Learning (ML) to revolutionize traffic volume estimation. Our cutting-edge solution leverages:

* Advanced computer vision techniques to analyze real-time camera feeds
* Machine learning algorithms to identify patterns and anomalies in traffic flow
* Integration with IoT sensors and real-time data feeds for unparalleled accuracy "With TrafficIntelligence, traffic managers can:
* Accurately estimate traffic volume in real-time
* Identify trends and patterns to inform traffic management decisions
* Optimize traffic signal timing and resource allocation
* Enhance traveler experience and reduce congestion

"Join us in transforming traffic management with AI-powered traffic volume estimation. Discover how TrafficIntelligence can help you make data-driven decisions and create smarter, more efficient transportation systems."

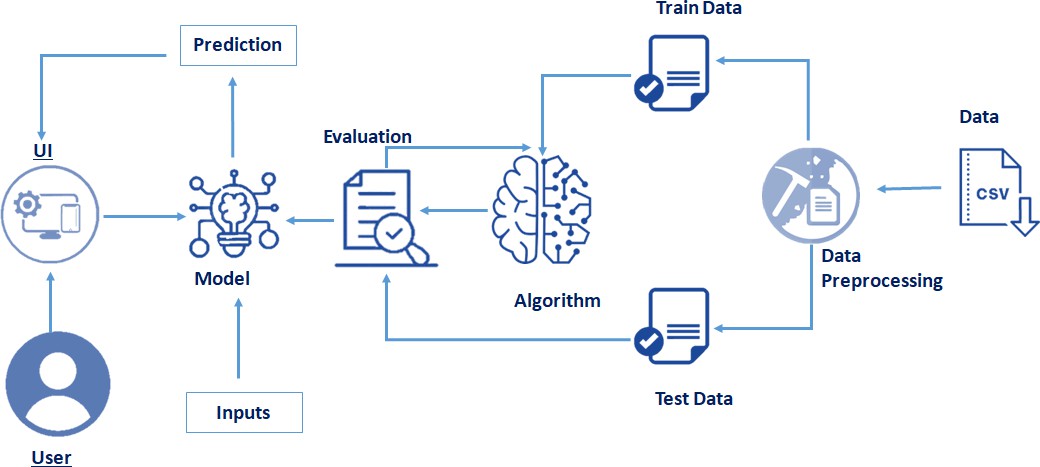
Would you like me to modify the introduction or provide further assistance?

**Scenario 1: Dynamic Traffic Management** TrafficTelligence enables dynamic traffic management by providing real-time traffic volume estimations. Transportation authorities can use this information to implement adaptive traffic control systems, adjust signal timings, and optimize lane configurations to reduce congestion and improve traffic flow.

**Scenario 2: Urban Development Planning** City planners and urban developers can leverage TrafficTelligence predictions to plan new infrastructure projects effectively. By understanding future traffic volumes, they can design road networks, public transit systems, and commercial zones that are optimized for traffic efficiency and accessibility.

**Scenario 3: Commuter Guidance and Navigation** Individual commuters and navigation apps can benefit from TrafficTelligence's accurate traffic volume estimations. Commuters can plan their routes intelligently, avoiding congested areas and selecting optimal travel times based on predicted traffic conditions. Navigation apps can provide real-time updates and alternative routes to improve overall travel experiences.

## Technical Architecture

****

**Project Objectives**

By the end of this project:

* You’ll be able to understand the problem to classify if it is a regression or a classification kind of problem.
* You will be able to know how to pre-process/clean the data using different data pre- processing techniques.
* You will able to analyze or get insights into data through visualization.
* Applying different algorithms according to a dataset and based on visualization.
* You will be able to know how to find the accuracy of the model.
* You will be able to know how to build a web application using the Flask framework.

## Project Flow

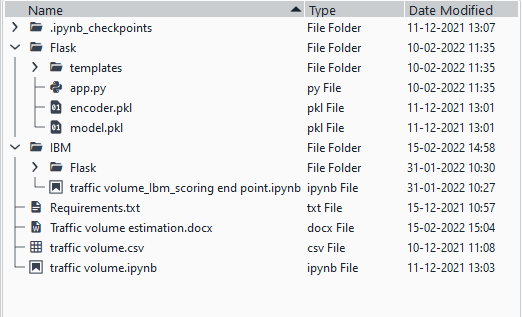
* + User interacts with the UI (User Interface) to enter the input values.
  + Entered input values are analyzed by the model which is integrated.
  + Once the model analyses the input the prediction is showcased on the UI.

To accomplish this, we have to complete all the activities and tasks listed below

* + Data Collection.
    - Collect the dataset or Create the dataset
  + Data Pre-processing.
    - Import the Libraries.
    - Importing the dataset.
    - Checking for Null Values.
    - Data Visualization.
    - Taking care of Missing Data.
    - Feature Scaling.
    - Splitting Data into Train and Test.
  + Model Building
    - Import the model building Libraries
    - Initializing the model
    - Training and testing the model
    - Evaluation of Model
    - Save the Model
  + Application Building
    - Create an HTML file
    - Build a Python Code
    - Run the App

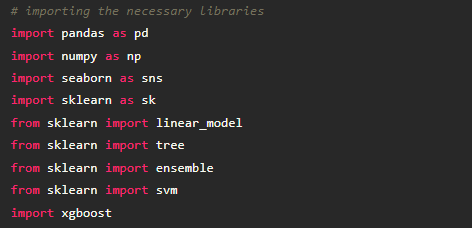
## Project Structure

Create a Project folder that contains files as shown below



* Flask files consist of template folder which has HTML pages, app.py file and .pkl files which are used for application building
* IBM folder has flask files and scoring endpoint.ipynb- model training code file.
* We need the model which is saved and the saved model in this content is Traffic volume

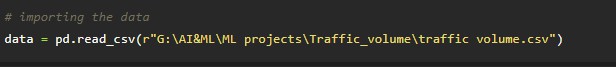
## Import Necessary Libraries

* It is important to import all the necessary libraries such as pandas, NumPy, matplotlib.
* **Numpy**- It is an open-source numerical Python library. It contains a multi-dimensional array and matrix data structures. It can be used to perform mathematical operations on arrays such as trigonometric, statistical, and algebraic routines.
* **Pandas-** It is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.
* **Seaborn**- Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
* **Matplotlib**- Visualisation with python. It is a comprehensive library for creating static,animated, and interactive visualizations in Python
* **Sklearn** – which contains all the modules required for model building.
* 

## Importing the Dataset

* You might have your data in .csv files, .excel files
* Let’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”)

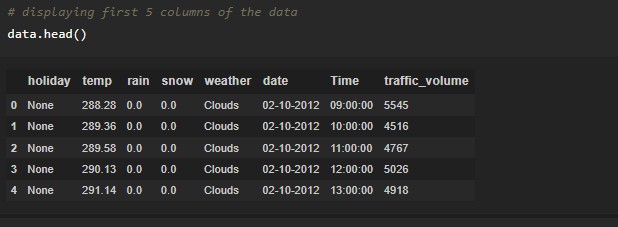
* 
* ​

**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 Columns
* Holiday - working day or holiday
* Temp- temperature of the day
* Rain and snow – whether it is raining or snowing on that day or not
* Weather = describes the weather conditions of the day
* Date and time = represents the exact date and time of the day
* Traffic 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

# Analyse the data

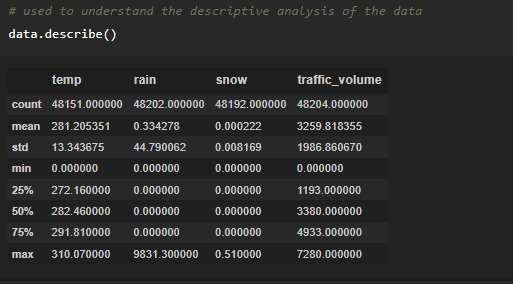
* **head()** method is used to return top n (5 by default) rows of a DataFrame or series.



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



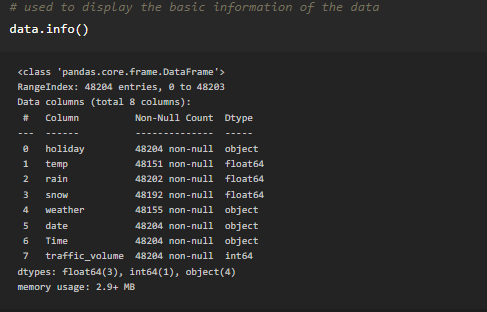
* The output is as shown below
* ​



* From the data, we infer that there are only decimal values and no categorical

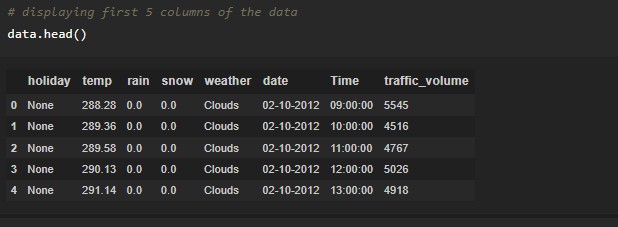
values.

**info()** gives information about the data - paste the image here.



# Analyse the data

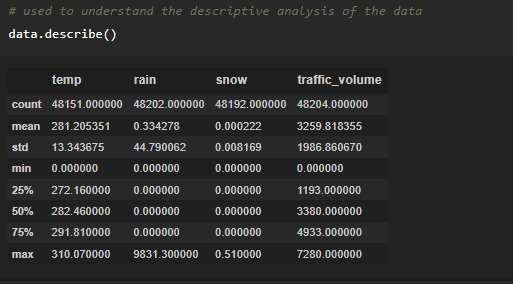
* **head()** method is used to return top n (5 by default) rows of a DataFrame or series.



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



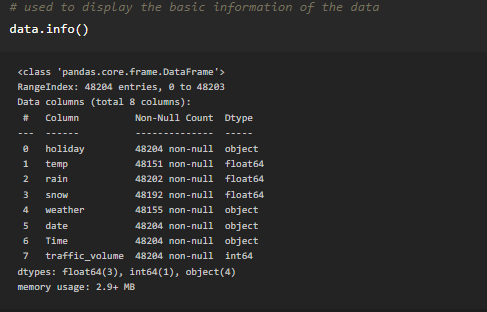
* The output is as shown below



* ​
* From the data, we infer that there are only decimal values and no categorical

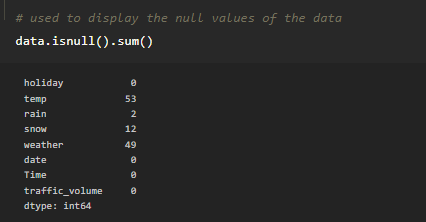
values.

**info()** gives information about the data - paste the image here.



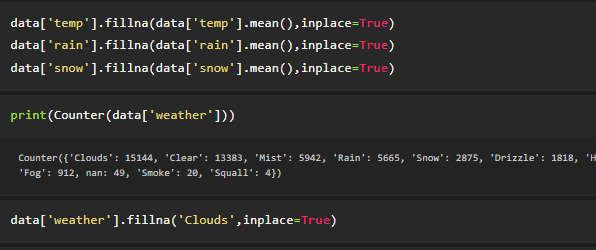
## Handling Missing Values

1. The Most important step in data pre-processing is dealing with missing data, the presence of missing data in the dataset can lead to low accuracy.
2. Check whether any null values are there or not. if it is present then the following can be done.



There are missing values in the dataset, we will fill the missing values in the columns.

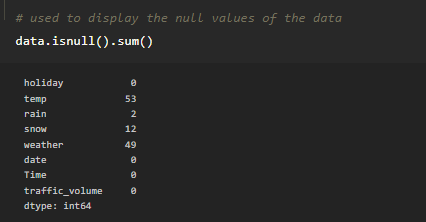
1. We are using mean and mode methods for filling the missing values
   * Columns such as temp, rain, and snow are the numeric columns, when there is a numeric column you should fill the missing values with the mean/median method. so here we are using the mean method to fill the missing values.
   * Weather column has a categorical data type, in such case missing data needs to be filled with the most repeated/ frequent value. Clouds are the most repeated value in the column, so imputing with clouds value.



pa

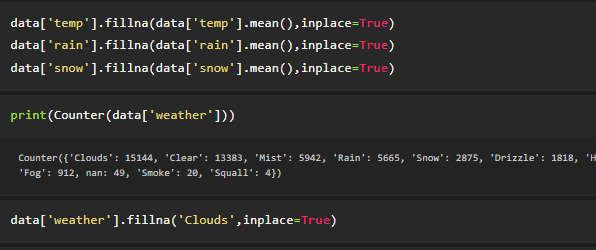
## Handling Missing Values

1. The Most important step in data pre-processing is dealing with missing data, the presence of missing data in the dataset can lead to low accuracy.
2. Check whether any null values are there or not. if it is present then the following can be done.



There are missing values in the dataset, we will fill the missing values in the columns.

1. We are using mean and mode methods for filling the missing values
   * Columns such as temp, rain, and snow are the numeric columns, when there is a numeric column you should fill the missing values with the mean/median method. so here we are using the mean method to fill the missing values.
   * Weather column has a categorical data type, in such case missing data needs to be filled with the most repeated/ frequent value. Clouds are the most repeated value in the column, so imputing with clouds value.



pa

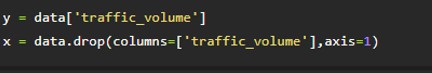
## Splitting the Dataset into Dependent and Independent variable

* In machine learning, the concept of the dependent variable (y) and independent variables(x) is important to understand. Here, the Dependent variable is nothing but output in dataset and the independent variable is all inputs in the dataset.
* With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

To read the columns, we will use iloc of pandas (used to fix the indexes for selection) which takes two parameters — [row selection, column selection].

Let’s split our dataset into independent and dependent variables. y = data[traffic\_volume] - independent

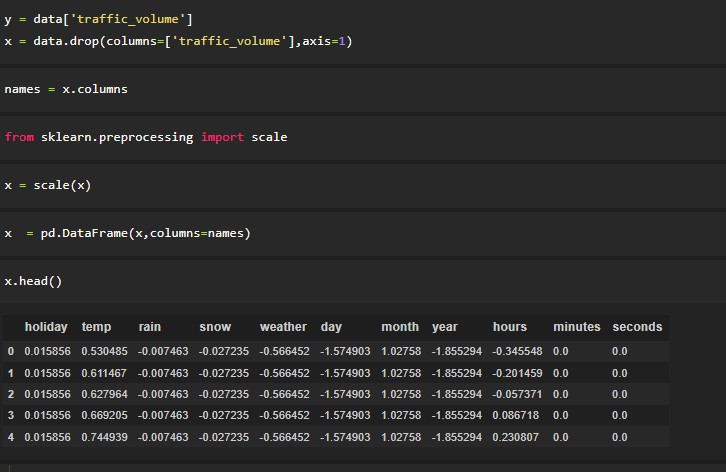
x = data.drop(traffic\_volume,axis=1)



## Feature Scaling

There is a huge disparity between the x values so let us use feature scaling.

Feature scaling is a method used to normalize the range of independent variables or features of data.



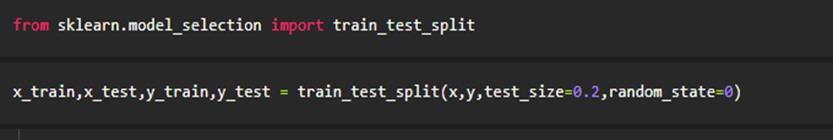
* + After scaling the data will be converted into an array form
  + Loading the feature names before scaling and converting them back to data frame after standard scaling is applied

## Splitting the data into Train and Test

When you are working on a model and you want to train it, you obviously have a dataset. But after training, we have to test the model on some test datasets. For this, you will a dataset which is different from the training set you used earlier. But it might not always be possible to have so much data during the development phase. In such cases, the solution is to split the dataset into two sets, one for training and the other for testing.

* + The train-test split is a technique for evaluating the performance of a machine learning algorithm.
  + Train Dataset: Used to fit the machine learning model.
  + Test Dataset: Used to evaluate the fit machine learning model.
  + In general you can allocate 80% of the dataset to the training set and the remaining 20% to test.
  + Now split our dataset into train set and test using train\_test\_split class from sci-kit learn library.

from sklearn import model\_selection x\_train,x\_test,y\_train,y\_test=model\_selection.train\_test\_split(x,y,test\_size=0.2,random\_ state =0)



## Model Building

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

## Training and Testing the Model

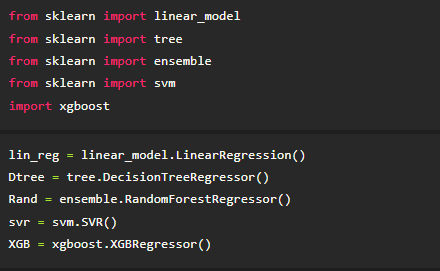
* + Once after splitting the data into train and test, the data should be fed to an algorithm to build a model.
  + There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose 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.KNN

5.svm

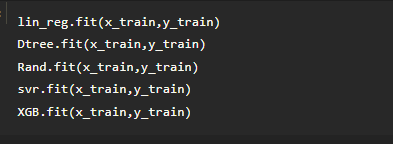
5.xgboost

Steps in Building the model:-

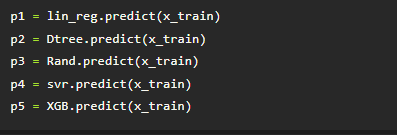
**Initialize the model -**

****

**Fit the models with x\_train and y\_train -**

****

**Predict the y\_train values and calculate the accuracy -**

****

We’re going to use the x-train and y-train obtained above in the train\_test\_split section to train our Random forest regression model. We’re using the fit method and passing the parameters as shown below.

We are using the algorithm from Scikit learn library to build the model as shown below,

Once the model is trained, it’s ready to make predictions. We can use the predict method on the model and pass x\_test as a parameter to get the output as y\_pred.

Notice that the prediction output is an array of real numbers corresponding to the input array.

## Model Evaluation

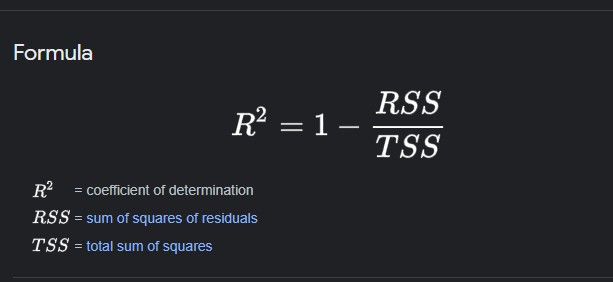
After training the model, the model should be tested by using the test data which is been separated while splitting the data for checking the functionality of the model.

## Regression Evaluation Metrics:

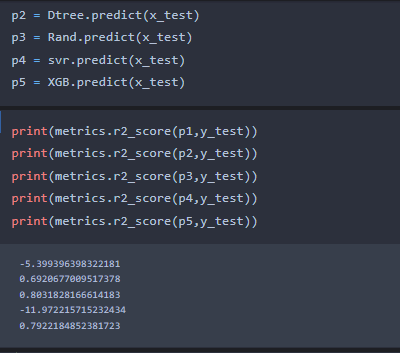
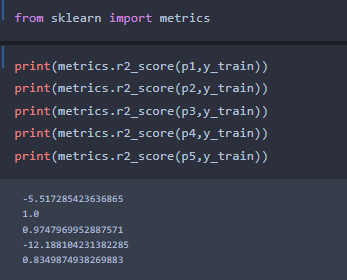
These model evaluation techniques are used to find out the accuracy of models built in the Regression type of machine learning models. We have three types of evaluation methods.

* + R-square\_score
  + RMSE – root mean squared error

## 1. R-squared \_score -

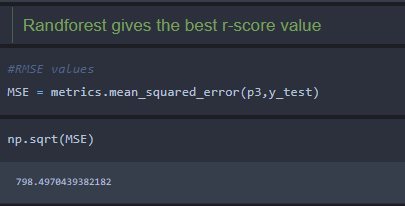


It is the ratio of the number of correct predictions to the total number of input samples.

Calculating the r2 score value using for all the models.

* After considering both r squared values of test and train we concluded that random forest regressor is giving the better value, it is able to explain the 97% of the data in train values.
* Random forest gives the best r2-score, so we can select this model.

## RMSE –Root Mean Square Error

****

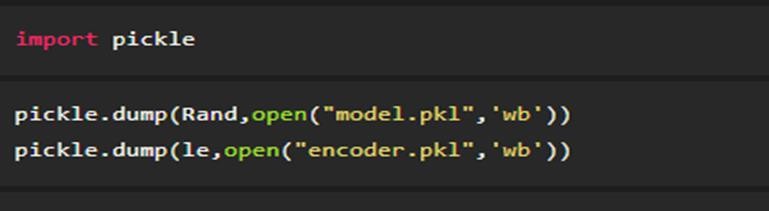
RMSE value for Random forest is very less when compared with other models, so saving the Random forest model and deploying using the following process

## Save the Model

After building the model we have to save the model.

Pickle in Python is primarily used in serializing and deserializing a Python object structure. In other words, it's the process of converting a Python object into a byte stream to store it in a file/database, maintain program state across sessions or transport data over the network. wb indicates write method and rd indicates read method.

This is done by the below code



## Application Building

In this section, we will be building a web application that is integrated into the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks

* Building HTML Pages
* Building server-side script

## Application Building

In this section, we will be building a web application that is integrated into the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks

* Building HTML Pages
* Building server-side script

## Main Python Script

Let us build an app.py flask file which is a web framework written in python for server-side scripting. Let’s see step by step procedure for building the backend application.

In order to develop web API with respect to our model, we basically use the Flask framework which is written in python.

Line 1-9 We are importing necessary libraries like Flask to host our model request Line 12 Initialise the Flask application

Line 13 Loading the model using pickle Line 16 Routes the API URL

Line 18 Rendering the template. This helps to redirect to the home page. In this home page,we give our input and ask the model to predict

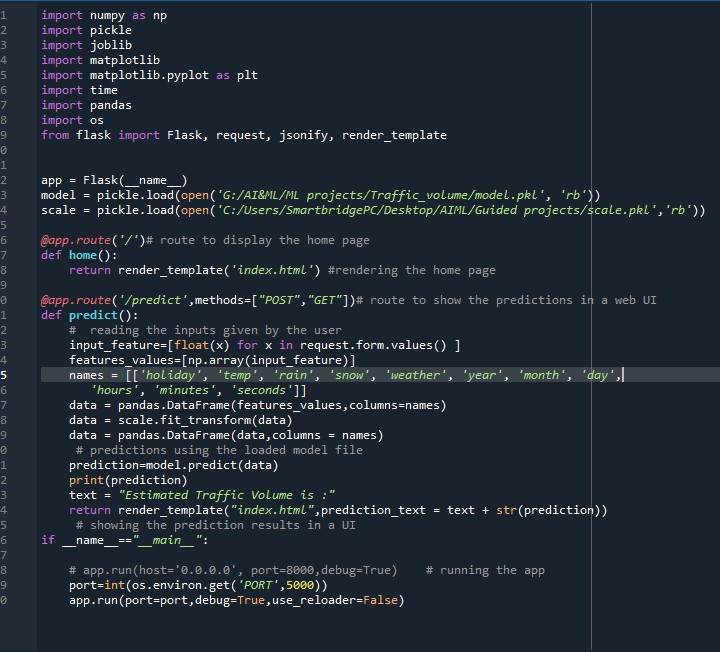
In line 23 we are taking the inputs from the form

Line 28 Feature Scaling the inputs

Line 31 Predicting the values given by the user

Line 32-35 if the output is false render no chance template If the output is True render chance template

Line 36 The value of name is set to main when the module run as the main program otherwise it is set to the name of the module .

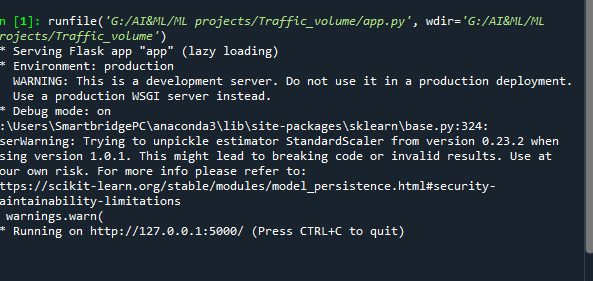


## Run the App

Open anaconda prompt from the start menu

* Navigate to the folder where your python script is.
* Now type the “python app.py” command

Navigate to the localhost where you can view your web page, Then it will run on **local host:5000**



## Output

* Copy the HTTP link and paste it in google link tab, it will display the form page
* Enter the values as per the form and click on predict button
* It will redirect to the page based on prediction output
* The output will be displayed in the prediction text as Estimated Traffic volume is in units.



Conclusion: In conclusion, TrafficIntelligence's AI-powered traffic estimation solution represents a paradigm shift in traffic management. By harnessing the power of machine learning, computer vision, and real-time data analytics, we've created a system that provides unparalleled accuracy, efficiency, and scalability.

"With TrafficIntelligence, traffic managers can:

* Make data-driven decisions with confidence
* Optimize traffic signal timing and resource allocation
* Reduce congestion, travel times, and emissions
* Enhance traveler experience and safety

"As we continue to push the boundaries of innovation, we envision a future where traffic management is:

* Predictive: anticipating and mitigating congestion before it occurs
* Adaptive: adjusting to changing traffic conditions in real-time
* Integrated: seamlessly connecting transportation modes and infrastructure