

Weekly progress report

Name: Ch.Shiva kumar

Domain: Data science and machine Learning

Date of submission:4/4/2024

Week Ending:02

Steps in the project execution:

LOAD AND CLEAN DATA:

1)Download the data sets for source link

<https://drive.google.com/file/d/1y61cDyuO9Zrp1fSchWcAmCk0B6SMx7X/view?usp=sharing>

2)We can clean data by using the Two methods:

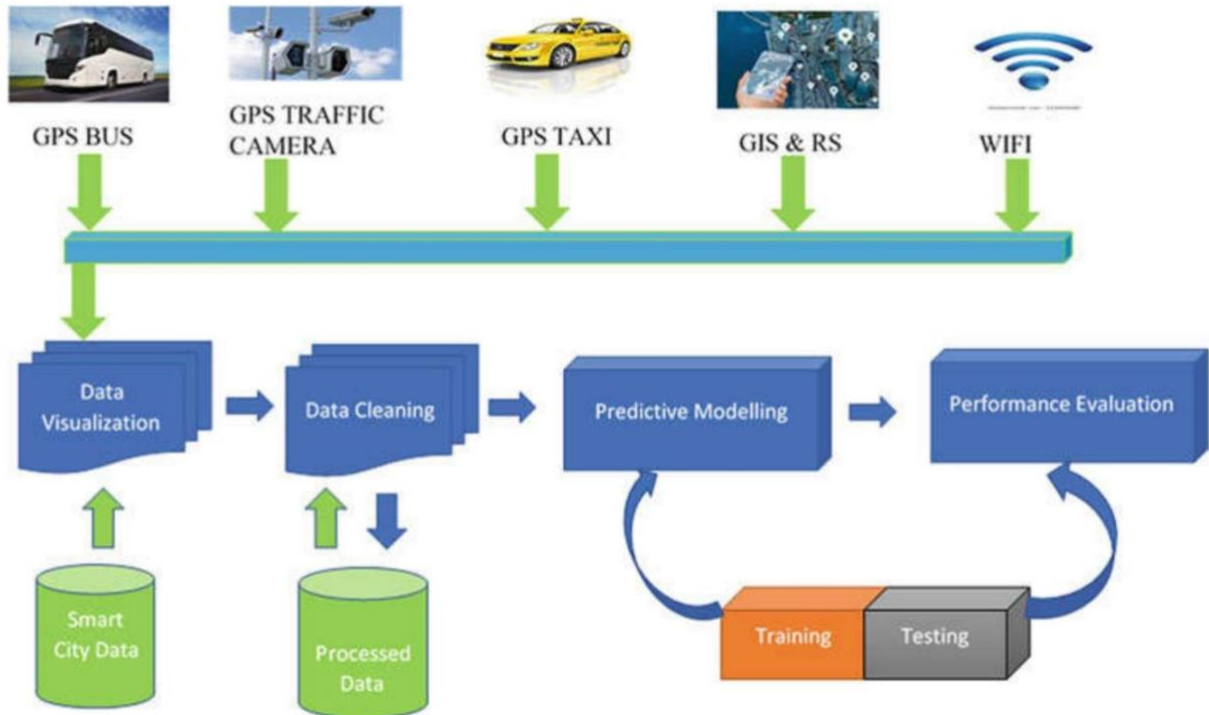
➤ By Using the Power-Bi desktop (or) Tableau software applications

- ✓ Open the Power Bi First
- ✓ Let take the data from the Option 'Get data'
- ✓ Next click on from the Common data sources As "Excel workbook"
- ✓ Select the file '2_Ecommerce Sales Data Analysis Excel.xlsx' from the required path
- ✓ Then Select All the Tables in the Dataset LOAD the data field
- ✓ Select the discount column and change the formula as "percentage"
- ✓ Delete the column field "Customer Name"
- ✓ Transform the data when completed of the above steps
- ✓ Choose Columns as Year,Order Date,Ship Date,Shipment days,Ship Mode,Segment,country,city,state,Region,Category,Sub_category,Sales,Quantity, Discount,Profit
- ✓ Change the Data fields as per Requirements
- ✓ Then Close & Apply
- ✓ **NOTE:-** If you want to know better about the dataset then perform the visualization by using different methods like:bar graph,pie-chart with to various locations and for different symbols in the traffic places to make the better smart City prediction

➤ By Using Python Modules

- ✓ import or load the data set by using python pandas module
- ✓ Syntax: variableName =pandas.read_csv("file location")
- ✓ import the modules numpy, pandas, matplotlib.pyplot
- ✓ Then load the training dataset and clean by below methods for remove null values, duplicate values and zero values
 - Variable_name=pandas.dropna(inplace=True)
 - Here Variable_name is loaded by the python

- `Variabl_name.drop_duplicates(inplace=True)`
- ✓ Then we can remove the duplicate values from the dataset.
- ✓ Now we can train the data model with this trained dataset
- ✓ Next we can take the testing dataset to test the model
- ✓ Before testing we need to clear the duplicate values from the testing dataset as above syntax like the training dataset



Block diagram for Forecasting of smart City traffic patterns

Apply operations by CSV file:

- 1) `var.head()`
 - Display the some top of the rows
- 2) `var.tail()`
 - Display the bottom of the rows
- 3) `var.info()`
 - Give the information of dataset
- 4) `var.describe`
 - Give the discription of dataset
- 5) `var.columns`
 - Give the all columns in dataset