**DRIVER DEMAND PREDICTION**

A Course End Project Submitted in Fulfillment of the Requirements

for the Course of

**A7512- MACHINE LEARNING**

In

# Department of Computer Science and Engineering

**Submitted By**

|  |  |
| --- | --- |
| **Name of the Student** | **Roll Number** |
| B Bhanulatha | 21881A05E1 |
| K.Krishna Varshita Reddy | 21881A05F8 |
| D.Abhi Ram | 21881A05E5 |

|  |
| --- |
| **CASE STUDY REPORT** |

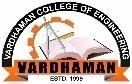
**VARDHAMAN COLLEGE OF ENGINEERING**

**(AUTONOMOUS)**

Affiliated to **JNTUH**, Approved by **AICTE**, Accredited by **NAAC** with **A++** Grade, **ISO 9001:2015** Certified

Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

**January 2024**

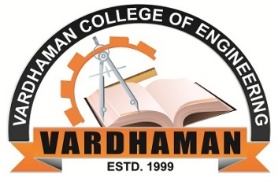
**VARDHAMAN COLLEGE OF ENGINEERING**

**(AUTONOMOUS)**

Affiliated to **JNTUH**, Approved by **AICTE**, Accredited by **NAAC** with **A++** Grade, **ISO 9001:2015** Certified

Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**CERTIFICATE**

Certified that this is a bonafide record of the course end project work entitled, **“*DRIVER DEMAND PREDICTION* ”**, done by, **B Bhanulatha (21881A05E1),K. Krishna Varshita Reddy(21881A05F8),D.Abhi Ram(21881A05E5)** submitted to the faculty of **Computer Science and Engineering**, in partial fulfillment of the requirements for the course of **Machine Learning** during the year 2023-2024 (V Semester).

Semester End Examination held on ……………………………………………**­­­­­­­­­­**

**Course Instructor: Head of Department:**

**Mr. A.Bhanu Prasad Dr. Ramesh Karnati**

**(Assistant Professor) (HOD, CSE)**

**List of Contents**

***page***

**CHAPTER 1 Problem Statement 4**

**CHAPTER 2 Description 5-6**

**CHAPTER 3 Code 7-8**

**CHAPTER 4 Result 9-10**

**CHAPTER 5 Conclusion 11 PROBLEM STATEMENT:**

The advent of ride-sharing platforms has revolutionized urban transportation, providing a convenient and efficient means of travel. However, one of the critical challenges faced by ride-sharing companies is the efficient allocation of drivers to meet fluctuating demand. Balancing the supply of drivers with the dynamic nature of ride requests is essential for optimizing operational efficiency, reducing passenger wait times, and maximizing overall user satisfaction.

The primary problem lies in accurately predicting and forecasting driver demand across different times, locations, and events. The lack of a reliable and real-time prediction system often leads to issues such as under-supply during peak hours or oversupply during low-demand periods. These inefficiencies can result in longer wait times for passengers, decreased driver earnings, and overall negative experiences for both drivers and riders. To address this challenge, it is imperative to develop a robust predictive modeling system that takes into account various factors influencing driver demand. Factors may include time of day, day of the week, weather conditions, special events, holidays, and geographic locations. The model should be capable of analyzing historical ride data to identify patterns, trends, and seasonality, and adapt in real-time to sudden changes in demand. Additionally, the model must consider external factors such as citywide events, traffic conditions, and public transportation disruptions, as they significantly impact ride-sharing demand. Integrating data from multiple sources and employing machine learning algorithms will be crucial in achieving accurate and timely predictions. The successful implementation of a driver demand prediction system will not only enhance the overall efficiency of ride-sharing services but also contribute to improved driver satisfaction, increased earnings, and a better experience for passengers. This project aims to develop an innovative solution that addresses the complex and dynamic nature of driver demand, ultimately optimizing resource allocation and improving the overall effectiveness of the ride-sharing ecosystem.

Top of Form

# **DESCRIPTION:**

Driver demand forecasting involves the use of data analysis and predictive modeling techniques to estimate the future demand for drivers in the ride-sharing or transportation industry. The objective is to anticipate fluctuations in the demand for rides across different time periods, locations, and scenarios. By accurately predicting the need for drivers, ride-sharing companies can optimize their operations, enhance customer satisfaction, and improve overall efficiency.

Driver demand prediction is the process of using historical data to forecast future driver requests in a particular area. Managers may pre-allocate driver resources in cities with the aid of accurate and real-time demand forecasting, which helps drivers find clients more quickly and cuts down on passenger waiting times. This project is aimed to choose the best model in predicting the driver demand where we use various Machine learning techniques such as regression analysis and time series forecasting. Various baseline models, including moving averages (simple, weighted, and exponential), linear regression with grid search, random forest regressor with random search, and XGBoost regressor with random search are used. We find out which model is more suitable in predicting the output using the metrics we obtain.

The ability to predict taxi demand in advance can help significantly to alleviate the problem of inadequate taxi supply. By using an accurate time-series forecasting framework, demand for the next time interval can be predicted for a given area. If the predicted demand in that area decreases while the supply is high, we can safely conclude that many taxis will be running vacant because they have not been reallocated to appropriate areas (the supply can be predicted from pings generated by taxis during their journeys). These idle taxis should be redirected immediately to areas with high unmet demand, balancing the flow between demand and supply and truncating the overall idle driving time and avoidable fuel consumption. Conversely, oversupply can lead to traffic congestion. Hence, it is necessary to understand the variable needs of the population in both time and space, which can be achieved only through a robust demand-prediction model.Accurate demand forecasting is the cornerstone of successful retail operations. It influences inventory management, shapes customer expectations, and sets the stage for a competitive edge. It uses historical sales data and sophisticated analytics to provide estimates about future customer demand. It allows businesses to make informed decisions in planning production, managing inventory, improving supply chain efficiency, and creating budget and revenue projections.Demand helps fuel profits and the economy. That's why it's an important concept. Demand is closely related to the concept of supply. While consumers try to pay the lowest prices they can for goods and services, suppliers try to maximize profits.

The Root Mean Squared Error (RMSE) is one of the two main performance indicators for a regression model. It measures the average difference between values predicted by a model and the actual values. It provides an estimation of how well the model is able to predict the target value (accuracy).

Mean Absolute Error (MAE) is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a [regression model](https://deepchecks.com/glossary/regression/).

The MAE loss function formula:

* **MAE = (1/n) Σ(i=1 to n) |y\_i – ŷ\_i|**
* n is the number of observations in the dataset.
* y\_i is the true value.
* ŷ\_i is the predicted value.

The MAE is a linear score, meaning all individual differences contribute equally to the mean. It provides an estimate of the size of the inaccuracy, but not its direction (e.g., over or under-prediction).

Mean absolute percentage error (MAPE) is a metric that defines the accuracy of a forecasting method. It represents the average of the absolute percentage errors of each entry in a dataset to calculate how accurate the forecasted quantities were in comparison with the actual quantities.

**Code:**

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error

from sklearn.preprocessing import StandardScaler

data = pd.read\_csv('driver-data[1].csv')

y = data['mean\_dist\_day'] # Target variable

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

model = LinearRegression()

model.fit(X\_train, y\_train)

y\_pred = model.predict(X\_test)

mse = mean\_squared\_error(y\_test, y\_pred)

mae = mean\_absolute\_error(y\_test, y\_pred)

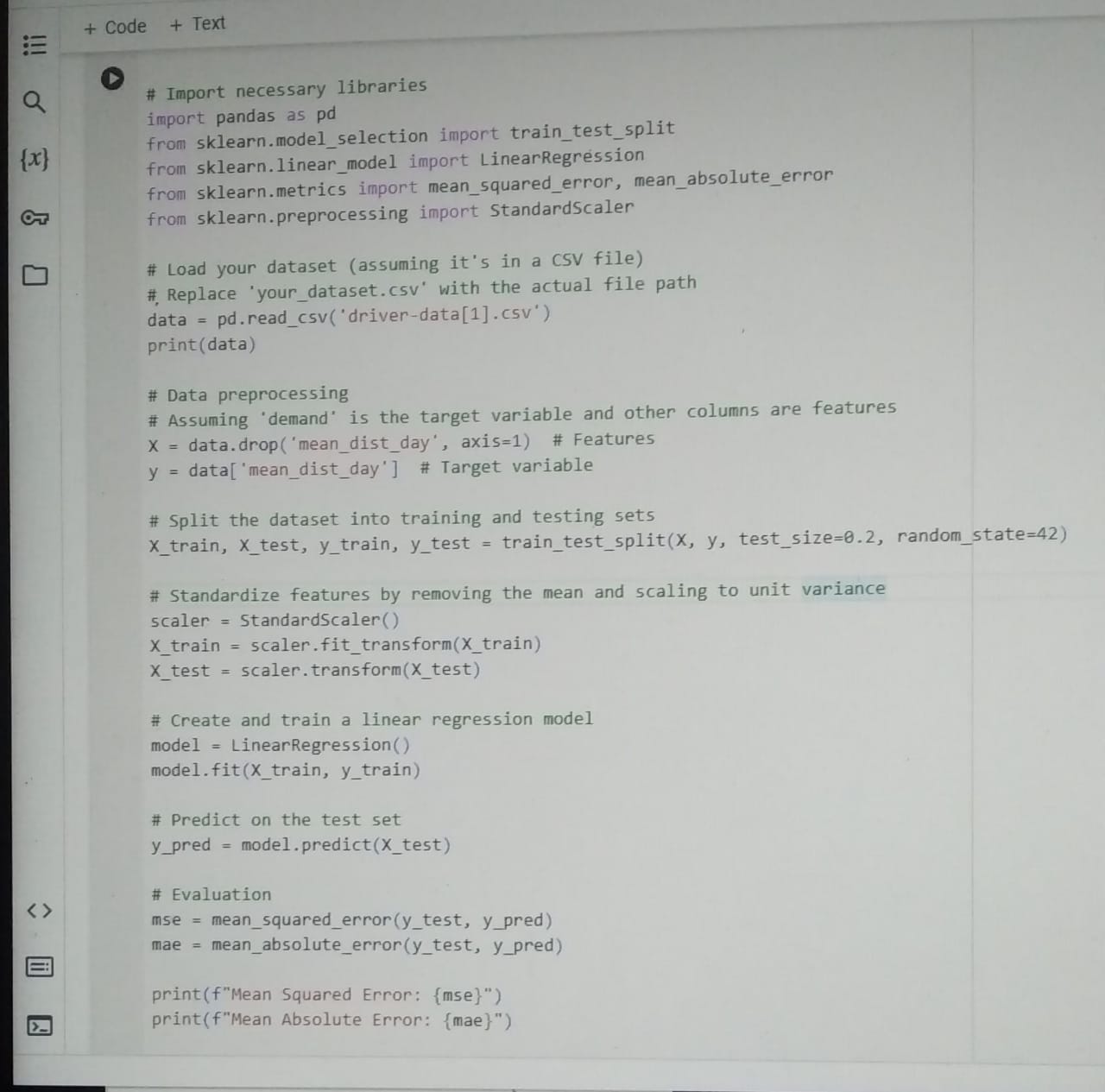
print(f"Mean Squared Error: {mse}")

print(f"Mean Absolute Error: {mae}")

**Code Explanation**:

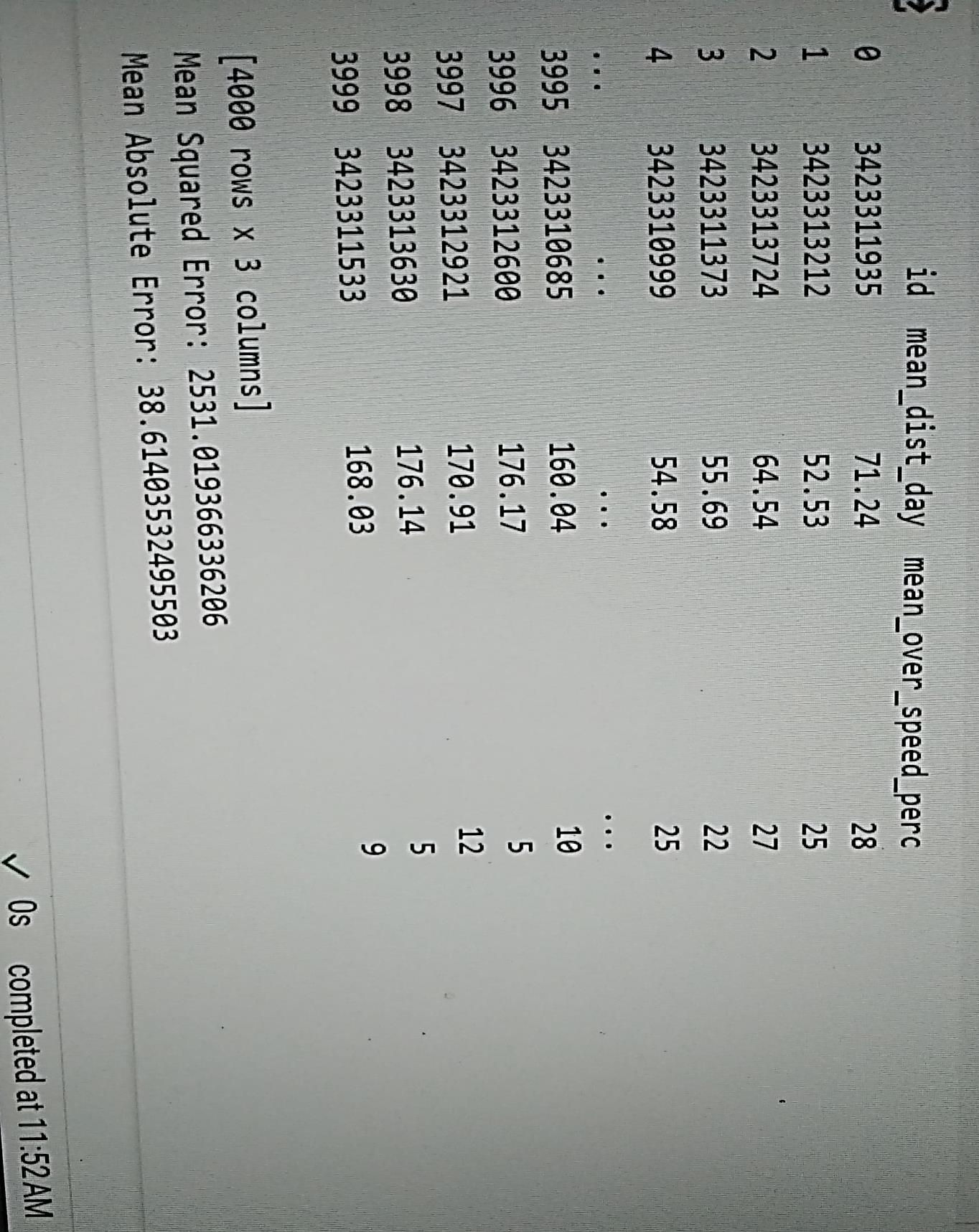
sklearn. model\_selection . train\_test\_split. Split arrays or matrices into random train and test subsets.

Here, train\_test\_split() and validation\_split do the same thing in splitting the data into train/test sets. The difference is that validation\_split does this in a way that makes the sets inaccessible after training, but evaluates them during training.The optimal split ratio depends on various factors. The rough standard for train-validation-test splits is 60-80% training data, 10-20% validation data, and 10-20% test data.To split the data,Create the Data Set or create a dataframe using Pandas. Shuffle data frame using sample function of Pandas. Select the ratio to split the data frame into test and train sets.To choose a train test split ratio,A commonly used ratio is 80:20, which means 80% of the data is for training and 20% for testing. Other ratios such as 70:30, 60:40, and even 50:50 are also used in practice.To perform a train-test split, use libraries like scikit-learn in Python. Import the `train\_test\_split` function, specify the dataset, and set the test size (e.g., 20%). This function randomly divides the data into training and testing sets, preserving the distribution of classes or outcomes



­

**Output:**



**Conclusion:**

Predicting driver demand is a crucial aspect of optimizing transportation systems and ensuring efficient and responsive services. Based on available data and advanced predictive analytics, several key conclusions can be drawn regarding driver demand prediction:

In conclusion, driver demand prediction is a crucial component of modern transportation systems. Leveraging data, machine learning models, and real-time adjustments can lead to more efficient and responsive services, providing benefits to both service providers and end-users. However, ongoing research and development are essential to address the challenges and uncertainties associated with predicting dynamic transportation demands accurately.