****

**Assessment Report**

on

**“Traffic Volume Prediction”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSEAI**

By

Yash Bansal (202401100300283)

**Under the supervision of**

“Prof. Abhishek Shukla”

**KIET Group of Institutions, Ghaziabad**

Affiliated to

**Dr. A.P.J. Abdul Kalam Technical University, Lucknow**

(Formerly UPTU)

**May, 2025**

**Introduction**

**Problem Statement**:  
The goal of this project is to build a regression model that predicts traffic volume on interstate highways based on weather conditions and time-related features. This is useful for city planners, traffic management systems, and smart city applications to prevent traffic congestion and improve commute efficiency.

**Dataset Used**:  
The dataset is taken from Kaggle (via kagglehub), named **Metro Interstate Traffic Volume**, and contains hourly traffic volume data alongside weather information and timestamps.

**Motivation**:  
Predicting traffic volume using machine learning can help improve traffic flow, reduce pollution, and provide real-time solutions in smart cities.

**Methodology**

**Step 1: Data Collection**

* The dataset was downloaded using the KaggleHub API.
* It contains features like temperature, rain, snow, clouds, and datetime.

**Step 2: Data Preprocessing & Feature Engineering**

* Converted datetime into useful time features like hour, day\_of\_week, and month.
* Added is\_weekend as a binary feature (0 for weekdays, 1 for weekends).
* Removed irrelevant columns for this task.

**Step 3: Model Selection**

* Used a basic **Linear Regression** model for simplicity and interpretability.
* Split the data into 80% training and 20% testing sets.

**Step 4: Training and Evaluation**

* Trained the model on the training set.
* Evaluated using **Mean Squared Error (MSE)**.

**Step 5: Visualization**

* Created scatter plots to visualize actual vs predicted traffic volume.
* Created a correlation heatmap to show relationships between features.

**CODE:**

*import kagglehub*

*# Download latest version*

*path = kagglehub.dataset\_download("rgupta12/metro-interstate-traffic-volume")*

*import pandas as pd*

*import numpy as np*

*import seaborn as sns*

*import matplotlib.pyplot as plt*

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

*from sklearn.linear\_model import LinearRegression*

*from sklearn.metrics import mean\_squared\_error, r2\_score*

*csv\_path = f"{path}/Metro\_Interstate\_Traffic\_Volume.csv"*

*# Step 2: Load Data*

*df = pd.read\_csv(csv\_path)*

*print("Dataset shape:", df.shape)*

*print(df.head())*

*# Step 3: Feature Engineering*

*df['date\_time'] = pd.to\_datetime(df['date\_time'])*

*df['hour'] = df['date\_time'].dt.hour*

*df['day\_of\_week'] = df['date\_time'].dt.dayofweek*

*df['month'] = df['date\_time'].dt.month*

*df['is\_weekend'] = df['day\_of\_week'].apply(lambda x: 1 if x >= 5 else 0)*

*# Step 4: Select Features*

*features = ['temp', 'rain\_1h', 'snow\_1h', 'clouds\_all', 'hour', 'day\_of\_week', 'month', 'is\_weekend']*

*X = df[features]*

*y = df['traffic\_volume']*

*# Step 5: Split Dataset*

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

*# Step 6: Train Model*

*model = LinearRegression()*

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

*# Step 7: Predict*

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

*# Step 8: Evaluate*

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

*print(f"Mean Squared Error: {mse:.2f}")*

*# Step 9: Visualize Actual vs Predicted*

*plt.figure(figsize=(8,5))*

*plt.scatter(y\_test, y\_pred, alpha=0.3, color='blue')*

*plt.xlabel("Actual Traffic Volume")*

*plt.ylabel("Predicted Traffic Volume")*

*plt.title("Actual vs Predicted Traffic Volume")*

*plt.grid(True)*

*plt.show()*

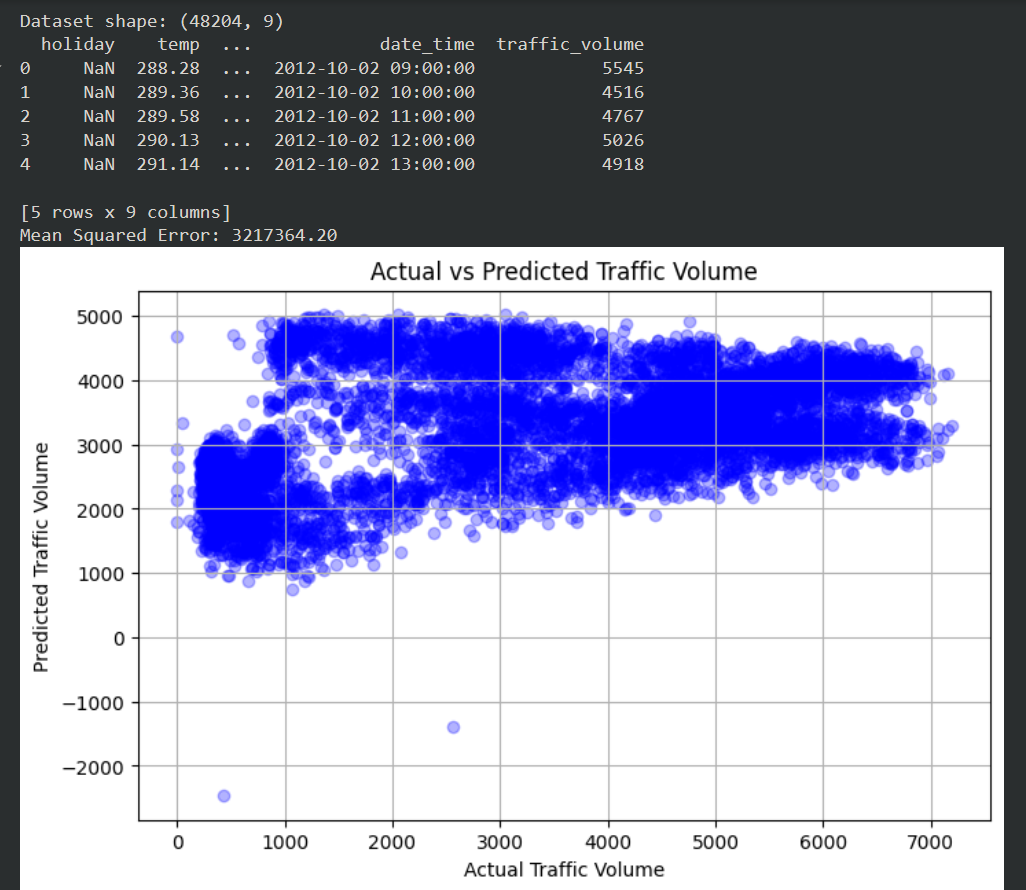
*# Step 10: Correlation Heatmap*

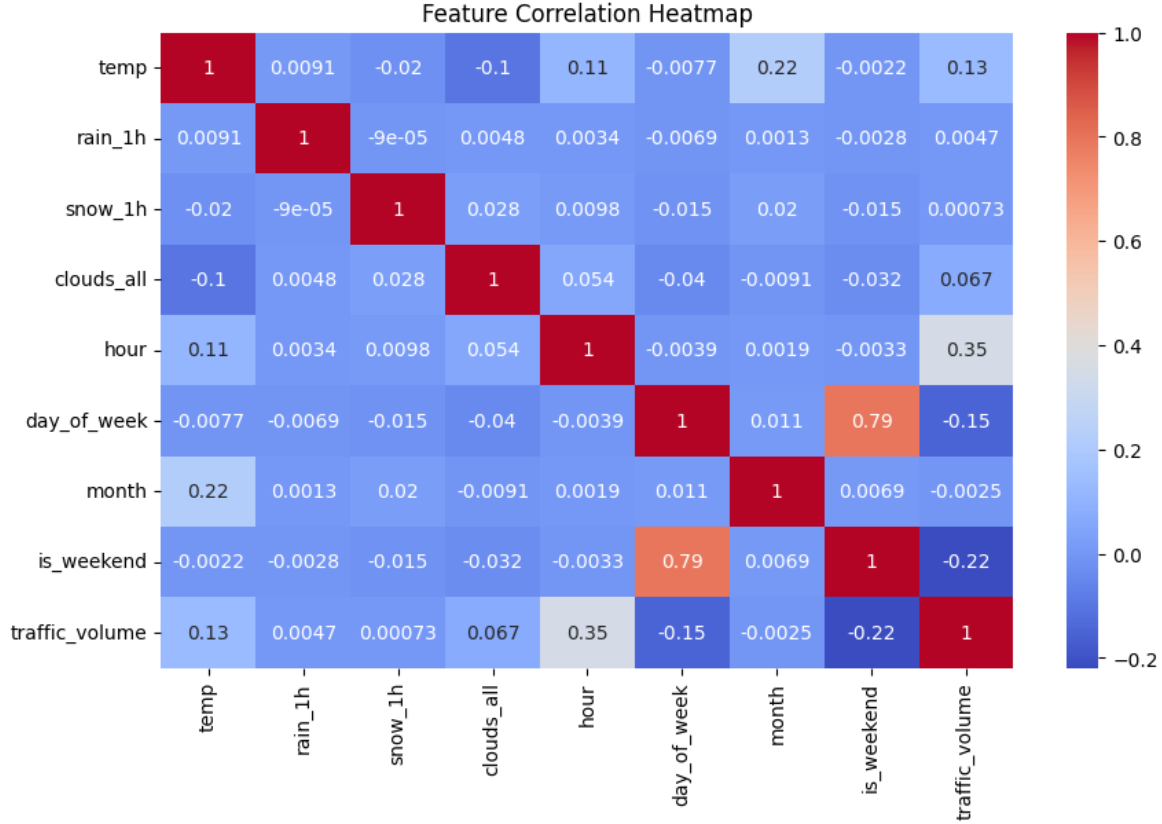
*plt.figure(figsize=(10,6))*

*sns.heatmap(df[features + ['traffic\_volume']].corr(), annot=True, cmap='coolwarm')*

*plt.title("Feature Correlation Heatmap")*

*plt.show()*

**Output/Result**

****

**References/Credits**

* **Dataset**: Metro Interstate Traffic Volume – Kaggle
* **Code Assistance**: Implemented using Python, scikit-learn, pandas, and matplotlib
* **IDE**: Jupyter Notebook / VS Code
* **API**: KaggleHub for downloading the dataset
* **Libraries**:
  + pandas – data handling
  + sklearn – model training
  + matplotlib and seaborn – visualization