Smart Traffic Flow Prediction and Congestion Detection System

1. Introduction

Traffic congestion is one of the major challenges in urban areas, leading to increased travel times, pollution, and fuel consumption. This project aims to predict traffic volume using historical traffic data, time-related features, and categorical variables. The model can help city planners and traffic management authorities identify rush hours, optimize signals, and improve road efficiency.

2. Objectives

The key objectives of this project are:

- • Analyze traffic patterns and identify peak hours.
- • Develop machine learning models to predict traffic volume.
- • Evaluate models using error metrics such as MAE, RMSE, and R^2.
- • Deploy a Google-Colab, based application for real-time predictions.

3. Dataset Description

The dataset used in this project (Traffic.csv) contains traffic count data collected across various times of the day. Key columns include:

- • Date – Day of the month
- • Time – Hourly intervals
- • Day of the week – Name of the weekday
- • Vehicle counts – Number of cars, bikes, buses, etc.
- • Total – Aggregate traffic volume
- • Traffic Situation – Category describing traffic level (High/Medium/Low)

4. Methodology

The project follows these steps:

1. Step 1 – Data Collection: Import traffic dataset.
2. Step 2 – Data Cleaning: Handle missing values, fix datatypes.
3. Step 3 – Feature Engineering: Extract hour, day, month, weekend indicators, and encode categorical features.
4. Step 4 – Exploratory Data Analysis (EDA): Visualize traffic by hour, day of week, and vehicle type.
5. Step 5 – Model Building: Train baseline Linear Regression, Random Forest, and XGBoost models.
6. Step 6 – Model Evaluation: Evaluate with MAE, RMSE, and R^2.

7. Step 7 – Deployment: Deploy Google_Colab, based interactive application.

5. Exploratory Data Analysis Results

From the analysis, the following insights were obtained:

- • Peak hours are observed between 6–8 AM and 4–7 PM.
- • Weekdays have significantly higher traffic compared to weekends.
- • Two-wheeler traffic is dominant compared to cars and buses.
- • Traffic Situation is mostly 'High' during rush hours.

6. Model Performance

Three models were tested and evaluated:

- • Linear Regression – Provided a simple baseline model.
- • Random Forest Regressor – Outperformed Linear Regression with better accuracy.
- • XGBoost Regressor – Achieved the best performance with optimized predictions.

Evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score.

7. Deployment

The project was deployed in Google Colab as an interactive notebook. Users can upload the traffic dataset, preprocess it, and train machine learning models directly within Colab. A custom prediction function allows users to input parameters such as hour, day of the week, month, and traffic situation to obtain real-time traffic volume predictions. Additionally, the notebook includes visualization dashboards to analyze peak hours, weekday vs weekend traffic, and congestion levels.

8. Conclusion

This project successfully demonstrates how machine learning can be applied to predict traffic volume. By the proper deployment techniques we can use such models in future situations can aid smart city initiatives, help reduce congestion, and optimize urban mobility. Future improvements can include adding weather data, real-time traffic feeds, and accident reports for even more accurate predictions.