Weekly Progress Report

Name: Afisar Alam

Domain: Data Science Machine Learning

Date of submission:27/12/2025

Week Ending: 01

Overview

This week, the primary focus was on understanding the Smart City Traffic Patterns dataset and setting up the end-to-end machine learning pipeline for traffic forecasting at four junctions.

I. Achievements:

1. Smart City Dataset Familiarization

Explored the train and test CSV files, understood
columns DateTime, Junction, Vehicles, and ID, and checked basic statistics and data quality.
Visualized vehicles distribution and junction-wise average traffic to understand overall patterns and seasonality.

2. Smart City Traffic Forecasting Project

Implemented data preprocessing in Python, including datetime conversion, feature engineering (hour, day of week, month, weekend flag, lag and rolling features).
Trained a RandomForestRegressor model with time-based train-validation split and evaluated performance using MAE and RMSE metrics on the validation set.

3. Learning Python / ML Libraries

Practiced using essential libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn for time-series style regression modelling.image.jpg
Gained confidence in building an end-to-end ML workflow from raw CSV data to final prediction file.

II. Challenges

1. Time-Series Feature Engineering

Faced difficulty in deciding which lag and rolling window sizes to use for capturing daily and hourly traffic patterns.
Addressed this by experimenting with multiple lags (1, 2, 3, 24) and rolling means (3 and 24 hours) and checking validation scores.

2. **Handling Warnings / Data Issues**
   Encountered library warnings (e.g., RMSE function deprecation) and confusion about the ID column presence in train vs test data.
   Resolved the issues by carefully inspecting data frame columns and switching to the recommended root_mean_squared_error function.

III. Learning Resources:

1. **Smart City Traffic Patterns (Kaggle)**
   Referred to the original dataset description and example notebooks on Kaggle to understand problem framing and common approaches**.**

2. **Python / ML Tutorials**
   Used online tutorials and documentation for Pandas and Scikit-learn to implement feature engineering, model training, and evaluation for regression problems**.**

Next Week's Goals:

1. **Model Enhancement**
   Try alternative models (e.g., Gradient Boosting / XGBoost) and perform basic hyperparameter tuning to further reduce validation error
   Analyze junction-wise performance and adjust features for under-performing junctions.

2. **Documentation and Visualization**
   Clean the Jupyter notebook, add clear markdown explanations and plots for each step, and start drafting the formal project report and README for GitHub.

Additional Comments:

This week provided a strong foundation in working with real-world traffic data and building a practical forecasting model that can support smart-city planning decisions during the internship.