UPSKILLS DATA SCIENCE AND MACHINE LEARNING INTERNSHIP

WEEK - 4

RAKSHA

GOVT. ENGINEERING COLLEGE, AJMER

UPSKILLS DATA SCIENCE AND MACHINE LEARNING INTERNSHIP

WEEK - 4

I would like to provide you with a progress report for my fourth week in the Upskills UCT Machine Learning and Data Science Internship. The following points highlight the key aspects of my activities and experiences:

Project Overview:

The Smart City Traffic Pattern ML project aims to analyze and predict traffic patterns in a smart city environment using machine learning techniques. By understanding and predicting traffic patterns, we can optimize traffic flow, improve transportation efficiency, and enhance overall urban mobility. This report provides an overview of the problem statement and discusses potential algorithms that can be employed in the project.

Problem Statement:

You are working with the government to transform your city into a smart city. The vision is to convert it into a digital and intelligent city to improve the efficiency of services for the citizens. One of the problems faced by the government is traffic. You are a data scientist working to manage the traffic of the city better and to provide input on infrastructure planning for the future. The government wants to implement a robust traffic system for the city by being prepared for traffic peaks. They want to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays, as well as on various other occasions during the year, differ from normal working days. This is important to take into account for your forecasting.

Now we done the basic study of the PS and our dataset and evaluated the following facts about the given dataset and what we have to submit in the final project report. So, we will work accordingly.

Data Dictionary

Variable	Description
ID	Unique ID
DateTime	Hourly Datetime Variable
Junction	Junction Type
Vehicles	Number of Vehicles (Target)

sample_submission.csv

Column Name	Description
ID	Unique ID
Vehicles	Number of Vehicles (Target)

Progress Summary: During Week 4, we focused on refining machine learning models, implementing real-time traffic monitoring, and enhancing the data preprocessing pipeline. The key activities and achievements are outlined below:

- **Refining Machine Learning Models:**

- Further refined the predictive models by incorporating additional features such as lag features generated from date and time information.
- Leveraged the temporal nature of traffic data to create lagged variables, capturing the historical patterns and trends.
- Incorporated these lag features into the machine learning models to improve their accuracy and predictive power.
    - **Data Preprocessing Enhancements:**
- Implemented feature scaling techniques to normalize the data and ensure fair comparisons across different features.
- Utilized both Min-Max scaling and Standardization scaling methods to handle different types of features.
- Applied Min-Max scaling to limit feature values between 0 and 1, preserving the data's original range.
- Used Standardization scaling to transform features into a standard normal distribution with mean 0 and standard deviation 1.
    - **Data Splitting:**
- Divided the dataset into training and testing (or validation) sets.
- Followed the common practice of allocating a certain percentage (e.g., 70% - 80%) of the data for training and the remaining portion for testing (or validation).
- Ensured that the temporal order of the data was maintained during the splitting process to mimic real-world scenarios.
    - **Data Reshaping:**
    - Reshaped the data to a format suitable for training machine learning models.

- **Prepared the input data in a way that the models could effectively learn the patterns and dependencies in the traffic data.**
- **Considered factors such as the number of time steps, lag features, and target variables to determine the appropriate reshaping strategy.**

Next Steps: Moving forward, the following tasks will be undertaken in Week 5:

- **Model Training and Evaluation:**

- Train the refined machine learning models using the preprocessed and reshaped data.
- Evaluate the performance of the models using appropriate metrics such as mean squared error, mean absolute error, or R-squared.
- Analyze the model results to assess their accuracy and determine if further optimization is required.

- **Real-Time Traffic Monitoring Enhancements:**

- Continuously monitor and fine-tune the real-time traffic monitoring system.
- Address any technical or operational challenges to ensure reliable and accurate data collection for analysis and forecasting.

- **Communication and Collaboration:**
- **Continuous Evaluation and Improvement:**

- Monitor the implemented traffic management strategies and assess their impact on traffic flow, congestion reduction, and overall citizen satisfaction.
- Gather feedback from citizens and stakeholders to identify areas for further improvements and implement necessary adjustments.

Challenges and Risks:

- Generating lag features from date and time information requires careful consideration of time intervals and potential seasonality effects.
- Scaling features using different techniques may introduce biases or distortions if not applied correctly or if the data distribution is not well understood.
- Ensuring the appropriate reshaping of the data is crucial for the models to effectively capture temporal dependencies and patterns.

Conclusion: During the fourth week of our smart city traffic prediction project, we successfully refined the data, enhanced the performed exploratory data analysis (EDA), and enhanced pre-processing steps, splitted the datasets and reshaped the data accordingly with our needs. These steps allowed us to gain insights into the dataset, identify patterns, and engineer meaningful features for our machine learning models and will also help us in future weeks to apply th algorithm finally. In the next phase, we will proceed with model selection and training based on this week progress.

Thanks and Regards

Raksha