UPSKILLS DATA SCIENCE AND MACHINE LEARNING INTERNSHIP

WEEK - 4

SNEHA AGRAWAL

MADHAV INSTITUTE OF TECHNOLOGY

& SCIENCE, GWALIOR

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 several important data preprocessing steps to enhance the accuracy and effectiveness of our traffic forecasting models. The key activities and achievements are outlined below:

- **Lag Feature Generation:**
- Incorporated lag features into our dataset to capture temporal dependencies in traffic patterns.

- Generated lag features based on date and time information, including previous hour, previous day, and previous week traffic data.
- This addition allowed our models to consider the historical traffic patterns and make more accurate predictions.
- **Scaling Features using Min-Max Scaler and Standard Scaler:**
- Applied feature scaling techniques to normalize the numerical features in our dataset.
- Utilized the Min-Max Scaler to scale the features to a specific range, typically between 0 and 1.
- Implemented the Standard Scaler to standardize the features by subtracting the mean and dividing by the standard deviation.
- Feature scaling ensured that all features were on a similar scale, avoiding dominance of certain features and enabling proper model convergence.
- **Splitting into Training and Testing (or Validation) Sets:**
- Divided our dataset into training and testing (or validation) sets to evaluate the performance of our models.
- The training set was used to train our predictive models, while the testing (or validation) set was used for model evaluation and fine-tuning.
- The splitting was done with consideration for maintaining the temporal order of the data, ensuring that earlier data was used for training and later data for evaluation.
- **Reshaping the Data:**
- Reshaped the data to meet the input requirements of our machine learning models.

- Utilized techniques such as sliding windows or time series forecasting structures to create sequences of data for model training.
- This reshaping allowed our models to learn from historical patterns and make predictions based on the sequential nature of the data.
- **Next Steps:** Moving forward, the following tasks will be undertaken in Week 5:
- **Model Training and Fine-Tuning:**
- Train our predictive models using the preprocessed and reshaped data.
- Evaluate the performance of the models on the testing (or validation) set and fine-tune the hyperparameters as necessary.
- Continuously monitor the models' performance and make adjustments to improve their accuracy and generalization ability.
- **Integration with Real-Time Traffic Monitoring:**
- Further enhance the real-time traffic monitoring system's capabilities by integrating it with the predictive models.
- Implement mechanisms to update the models with the latest real-time traffic data, allowing for adaptive and dynamic traffic forecasting.
- **Collaboration with Infrastructure Planning Teams:**
- Maintain close collaboration with infrastructure planning teams and stakeholders.
- Share the latest traffic forecasts, model insights, and evaluation results to inform infrastructure planning and decision-making processes.
- **Continuous Evaluation and Improvement:**

- Regularly evaluate the effectiveness of the implemented traffic management strategies based on real-time data and citizen feedback.
- Identify areas for improvement and make necessary adjustments to optimize traffic flow, minimize congestion, and enhance overall citizen satisfaction.

- **Challenges and Risks:**

- Generating lag features and reshaping the data can introduce computational and memory constraints, particularly for large datasets.
- Proper feature scaling is crucial to prevent biases in model training and ensure accurate predictions.
- Maintaining the integrity of the temporal order of data

- **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

Sneha Agrawal