UPSKILLS DATA SCIENCE AND MACHINE LEARNING INTERNSHIP

WEEK - 3

RAKSHA

GOVT. ENGINEERING COLLEGE, AJMER

UPSKILLS DATA SCIENCE AND MACHINE LEARNING INTERNSHIP

WEEK - 3

I would like to provide you with a progress report for my third 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 Of The Week:

The goal of this project is to develop a smart city traffic prediction system using machine learning techniques. In this report, we will cover the progress made during the third week of the project, which includes data loading, exploratory data analysis (EDA), and feature engineering.

2. Data Loading: During the third week, we focused on loading the necessary data for our smart city traffic prediction system. The data consists of historical traffic records collected from various sensors and sources in the city. We obtained the data from the city's transportation department, which provided us with a structured dataset in CSV format.

To load the data, we used the pandas library in Python. We read the CSV file into a pandas DataFrame, which allowed us to manipulate and analyze the data effectively. We verified the successful loading of the data by inspecting a sample of the records and ensuring that all the required columns were present.

3. Exploratory Data Analysis (EDA): After loading the data, we performed exploratory data analysis to gain insights into the dataset and understand its characteristics. The EDA process involved the following steps:

a) **Data Summary**: We generated summary statistics of the dataset, including measures of central tendency, dispersion, and distributions of different features. This allowed us to understand the range and variability of the data.

b) **Data Visualization**: We created various visualizations such as histograms, box plots, scatter plots, and time series plots to explore the relationships between different variables. These visualizations helped us identify patterns, outliers, and potential correlations within the data.

c) **Correlation Analysis**: We calculated the correlation matrix to quantify the relationships between different features. This analysis helped us identify highly correlated variables, which could potentially be used as informative features for our traffic prediction model.

d) **Data Quality Check**: We performed data quality checks to identify missing values, outliers, and inconsistencies in the dataset. We used techniques such as checking for null values, data imputation, and removing or correcting outliers to ensure the data's integrity and reliability.

4. Feature Engineering: Feature engineering is a crucial step in machine learning projects as it involves transforming raw data into meaningful features that can improve the predictive performance of the models. In this phase, we performed the following feature engineering tasks:

a) **Temporal Features**: We extracted useful temporal features from the timestamp column, such as hour of the day, day of the week, month, and year. These features can capture the periodicity and time-dependent patterns in traffic data.

b) **Categorical Encoding**: We encoded categorical variables using techniques like one-hot encoding or label encoding, depending on the nature of the variables. This conversion allowed us to represent categorical data numerically, making it suitable for machine learning algorithms.

c) **Feature Scaling**: We applied feature scaling techniques such as standardization or normalization to ensure that all the features were on a similar scale. This step is important because many machine learning algorithms perform better when the features are normalized.

d) **Feature Selection**: We used various feature selection techniques, such as correlation analysis, statistical tests, and domain knowledge, to identify the most relevant features for our traffic prediction model. This process helped us reduce the dimensionality of the dataset and focus on the most informative variables.

5. Conclusion: During the third week of our smart city traffic prediction project, we successfully loaded the data, performed exploratory data analysis (EDA), and conducted feature

engineering. These steps allowed us to gain insights into the dataset, identify patterns, and engineer meaningful features for our machine learning models. In the next phase, we will proceed with model selection and training based on this week progress.

Thanks and Regards

Raksha