
# ADVANCED TECHNIQUES IN DATA SCIENCE

# SEMESTER PROJECT

**Submitted To:** **Dr. Awais Hassan**

**Submitted By:** **Rimsha Shafique**

**Registration No.:** **2024-MSDS-03**

**Semester:** **1st**

**Department:** **Data Science**

***Date of submission:*** 25 December, 2024

# Problem Statement: *Predicting High-Demand Levels for Ride Services*

**Background**

Efficient resource allocation and demand management are critical for ride service providers like Uber. Predicting periods of high demand enables better decision-making, such as optimizing driver availability and pricing strategies, ensuring customer satisfaction, and maximizing revenue. An accurate prediction model can significantly enhance operational efficiency and service quality.

**Objective**

The primary goal of this project is to predict high-demand periods for ride services using key features such as the day of the week, time of the day, ride distance, and ride category. This involves training a machine learning model to classify demand levels as **High (1)** or **Low (0)** based on historical ride data.

**Scope and Significance**

This study aims to:

1. Develop a predictive model for demand levels using a Random Forest Classifier.
2. Identify the most important features influencing ride demand levels, including time, distance, day, and ride category.
3. Provide actionable insights for improving operational efficiency and strategic planning in ride services.

**Methodology**

1. **Data Aggregation**:
   - Group historical ride data by time periods to calculate ride counts.
   - Define high-demand thresholds using the 75th percentile of ride counts.
   - Label demand levels as "High" or "Low" based on the thresholds.
2. **Feature Selection**:

   Use features such as:

   - **Day**: Day of the week.
   - **Time**: Hour of the day.
   - **Miles**: Distance of the ride.
   - **Category**: Type of ride (e.g., Business, Personal).
3. **Model Development**:

- Preprocess data using one-hot encoding for categorical features and passthrough for numerical features.
- Train a Random Forest Classifier to predict demand levels.
- Evaluate model performance using metrics like precision, recall, and F1-score.

4. **Feature Importance Analysis**:
    - Extract and visualize the importance of each feature in predicting demand levels to gain actionable insights.

**Expected Outcomes**

- A machine learning model capable of accurately predicting high and low demand levels.
- A ranked list of features based on their importance in determining demand levels.
- Visualizations and insights that inform strategic decisions for resource allocation and service optimization.

**Applications**

- **Driver Allocation**: Deploying drivers effectively during high-demand periods.
- **Dynamic Pricing**: Implementing surge pricing strategies during peak hours.
- **Operational Planning**: Enhancing scheduling and resource management for optimal service delivery.