**Bike Rental Prediction Project**Course-end Project 1

**Introduction**

In recent years, bike-sharing systems have gained popularity, providing a convenient mode of transportation. These systems allow users to rent bikes from one location and return them to another seamlessly. Understanding and predicting daily bike rental counts based on environmental and seasonal factors is essential for effective operations and planning. This project employs machine learning techniques to analyze and predict bike rental counts.

**Objective**

The primary objective of this project is to predict the total daily bike rentals (cnt) using environmental and seasonal features. By doing so, we aim to assist the bike rental company in optimizing their services.

**Dataset Overview**

The dataset (day.csv) contains daily records of bike rentals. Below is a summary of the key attributes:

| **Variable** | **Description** |
| --- | --- |
| instant | Record index |
| dteday | Date |
| season | Season (1: Spring, 2: Summer, 3: Fall, 4: Winter) |
| yr | Year (0: 2011, 1: 2012) |
| mnth | Month (1 to 12) |
| holiday | Whether the day is a holiday (1: Yes, 0: No) |
| weekday | Day of the week |
| workingday | Whether the day is a working day (1: Yes, 0: No) |
| weathersit | Weather condition (1: Clear, 2: Mist, 3: Light rain, 4: Heavy rain) |
| temp | Normalized temperature (max value divided by 41) |
| atemp | Normalized feeling temperature (max value divided by 50) |
| hum | Normalized humidity (max value divided by 100) |
| windspeed | Normalized wind speed (max value divided by 67) |
| casual | Count of casual users |
| registered | Count of registered users |
| cnt | Total count of rental bikes (casual + registered) |

**Steps to Solve the Problem**

**1. Exploratory Data Analysis (EDA)**

* **Load and Inspect the Data**:
  + Load the dataset and check its structure, dimensions, and summary statistics.
* **Data Type Conversion**:
  + Convert relevant columns (e.g., season, yr, mnth, etc.) to appropriate data types (e.g., factors).
* **Missing Value Analysis**:
  + Check for missing values and handle them if necessary.

**2. Attribute Distribution and Trends**

* **Monthly Distribution**:
  + Plot the average bike rentals for each month.
* **Yearly Distribution**:
  + Analyze and visualize bike rental trends across years.
* **Outlier Analysis**:
  + Use boxplots to detect outliers in rental counts across different seasons.

**3. Train-Test Split**

* Split the dataset into training (70%) and testing (30%) subsets to evaluate model performance.

**4. Random Forest Model**

* Train a Random Forest regression model using relevant features (e.g., temp, atemp, hum, windspeed, etc.) to predict the target variable cnt.
* Optimize hyperparameters for better performance.

**5. Model Evaluation**

* Predict the rental counts on the test dataset.
* Evaluate the model’s performance using metrics such as Mean Squared Error (MSE).
* Plot actual vs. predicted rental counts to visualize the model’s accuracy.

**Results and Analysis**

* **EDA Insights**:
  + Seasonal and weather-related patterns significantly influence bike rentals.
  + High rentals are observed during summer and fall, with lower counts in winter.
* **Model Performance**:
  + The Random Forest model achieved a low MSE, indicating accurate predictions of rental counts.
  + Visualizations confirmed a strong correlation between predicted and actual values.

**Conclusion**

This project successfully demonstrates the use of machine learning techniques to predict daily bike rentals. By analyzing environmental and seasonal trends, the bike rental company can:

* Optimize bike availability.
* Enhance user experience.
* Plan for seasonal variations in demand.

**Future Scope**

* Explore advanced models such as Gradient Boosting or Neural Networks for improved accuracy.
* Include additional features such as population density or special events to enhance predictions.
* Implement real-time prediction systems for dynamic demand forecasting.

This project showcases the power of exploratory data analysis, visualization, and machine learning in solving real-world business problems effectively.