



Project 2: Bike Rental Prediction

Objectives

- Perform exploratory data analysis and visualize the data to understand the environmental and seasonal settings
- Predict bike rental counts based on environmental and seasonal settings with the help of a machine learning algorithm



Prerequisites

- Exploratory data analysis
- Data manipulation
- Data visualization
- R programming
- Machine learning



Industry Relevance



- **Exploratory data analysis:** It finds trends and patterns, or checks assumptions by analyzing data with visual tools
- **Data manipulation:** It organizes and changes information to make it more understandable.
- **Data visualization:** It represents data with the use of common graphs, plots, or charts.
- **R programming:** It is used for statistical analysis, graphics representation, and reporting.
- **Machine Learning:** It helps software programs in being more accurate at predicting outcomes without explicitly programming them to do so.

Problem Statement



In bike-sharing systems, the entire process from membership to rental and return has been automated. Using these systems, users can easily rent a bike from one location and return it to another. Hence, a bike rental company wants to understand and predict the number of bikes rented daily based on the environment and seasons.

Dataset Description



Variable

Description

- | | | |
|------------|---|---|
| instant | - | Record index |
| dteday | - | Date |
| season | - | Season (1: springer, 2: summer, 3: fall, 4: winter) |
| yr | - | Year (0: 2011, 1: 2012) |
| mnth | - | Month (1 to 12) |
| holiday | - | Weather day is a holiday or not |
| weekday | - | Day of the week |
| workingday | - | Working day (1: neither weekend nor holiday, 0: other days) |

Dataset Description



Variable	-	Description
weathersit	-	1: Clear, few clouds, partly cloudy, partly cloudy 2: Mist + cloudy, mist + broken clouds, mist + few clouds, mist 3: Light snow, light rain + thunderstorm + scattered clouds, light rain + scattered clouds 4: Heavy rain + ice pallets
temp	-	Normalized temperature in Celsius; The values are divided into 41 (max)
atemp	-	Normalized feeling temperature in Celsius; The values are divided into 50 (max)
hum	-	Normalized humidity; The values are divided into 100 (max)

Dataset Description



Variable	-	Description
windspeed	-	Normalized wind speed; The values are divided into 67 (max)
casual	-	Count of casual users
registered	-	Count of registered users
cnt	-	Count of total rental bikes including both casual and registered

Tasks to Perform

Perform the following tasks on the dataset provided using R:

1. Exploratory data analysis:

- Load the dataset and the relevant libraries
- Perform data type conversion of the attributes
- Carry out the missing value analysis

2. Attributes distribution and trends

- Plot monthly distribution of the total number of bikes rented
- Plot yearly distribution of the total number of bikes rented
- Plot boxplot for outliers' analysis



Tasks to Perform

Perform the following tasks on the dataset provided using R:

3. Split the dataset into train and test dataset
4. Create a model using the random forest algorithm
5. Predict the performance of the model on the test dataset



Project Outcome



- This project is designed to help understand how to perform exploratory data analysis, plot graphs, and predict using a machine learning algorithm.
- You should be able to analyze the dataset for this project to create a report. You will be able to use a machine learning algorithm and predict the bikes rented daily.

Submission Process



1. Complete the project in the Simplilearn lab
2. Complete each task listed in the problem statement
3. Take screenshots of the results for each question and the corresponding code
4. Save it as a document and submit using the assessment tab
5. Tap the "Submit" button (this will present you with three choices)
6. Attach three files and then click "Submit"

Note: Be sure to include screenshots of the output

Thank You