PRCP-1018-BikeRental

Problem Statement

Task 1:- Prepare a complete data analysis report on the given data.

Task 2:- Prediction of daily bike rental count based on the environmental and seasonal settings.

Dataset Description and link:

Bike sharing systems are a new version of traditional bike rentals where the whole process from membership, rental and return back has been automated. Through these systems, users are easily able to rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousands bicycles. Today, there exists great interest in these systems due to their important role in traffic, environmental and health issues. Apart from interesting real world applications of bike sharing systems, the characteristics of data being generated by these systems make them attractive for research. Opposed to other transport services such as bus or subway, the duration of travel, departure and arrival position is explicitly recorded in these systems. This feature turns the bike sharing system into a virtual sensor network that can be used for sensing mobility in the city. Hence, it is expected that most of the important events in the city could be detected via monitoring these data.

Link : https://d3ilbtxij3aepc.cloudfront.net/projects/CDS-Capstone-Projects/PRCP-1018-BikeRental.zip

Attribute Information:

Both the files hour.csv and day.csv have the following fields, except hr which is not available in day.csv

- instant: record index

- dteday : date

- season : season (1:winter, 2:spring, 3:summer, 4:fall)

- yr : year (0: 2011, 1:2012)

- mnth : month (1 to 12)

- hr : hour (0 to 23)

- holiday : weather day is holiday or not (extracted from [Web Link])

- weekday : day of the week

- workingday : if neither weekend nor holiday, it is 1; otherwise 0.

+ 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 Pellets + Thunderstorm + Mist, Snow + Fog

- temp : Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39 (only in hourly scale)

- atemp: Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50 (only in hourly scale)

- hum: Normalized humidity. The values are divided to 100 (max)

- windspeed: Normalized wind speed. The values are divided to 67 (max)

- casual: count of casual users

- registered: count of registered users

- cnt: count of total rental bikes including both casual and registered

Model Comparison Report

Create a report stating the performance of multiple models on this data and suggest the best model for production.

Report on Challenges faced

Create a report which should include challenges you faced on data and what technique was used with proper reasoning.

Note:-All above tasks need to be generated on a single jupyter notebook, which must be shared as final submission of the project.