

Re-Sit: Post Module Assessment

MODULE TITLE: Machine Learning
MODULE CODE: MCB008
SUBMISSION: 9th April 2021 by 12 Noon

Predict Number of Bike Rentals

A company that rents bicycles wants to predict the expected number of rentals in a given day, based on the season, day of the week, weather conditions, and so on.

The first step in this project is to explore the data that you will use to train a model. The goal of this exploration is to try to understand the relationships between its attributes; in particular, any apparent correlation between the *features* and the *label* your model will try to predict.

Your task is to build a predictive model using machine learning to predict the expected number of rentals in a given day.

Code

We are looking for you to show off your coding skills using Python.

Data

The data consists of the following columns:

instant: A unique row identifier

dteday: The date on which the data was observed - in this case, the data was collected daily; so, there's one row per date.

season: A numerically encoded value indicating the season (1:spring, 2:summer, 3:fall, 4:winter)

yr: The year of the study in which the observation was made (the study took place over two years - year 0 represents 2011, and year 1 represents 2012)

mnth: The calendar month in which the observation was made (1:January ... 12:December)

holiday: A binary value indicating whether or not the observation was made on a public holiday)

weekday: The day of the week on which the observation was made (0:Sunday ... 6:Saturday)

workingday: A binary value indicating whether or not the day is a working day (not a weekend or holiday)

weathersit: A categorical value indicating the weather situation (1:clear, 2:mist/cloud, 3:light rain/snow, 4:heavy rain/hail/snow/fog)

temp: The temperature in celsius (normalized)

atemp: The apparent ("feels-like") temperature in celsius (normalized)

hum: The humidity level (normalized)

windspeed: The windspeed (normalized)

rentals: The number of bicycle rentals recorded.

The files in this folder: **daily-bike-share.csv**

Tasks:

1. Import the data from **daily-bike-share.csv** and view the first few rows. (Hint- Use Python Pandas library)
2. Create Histogram to visualise data for feature 'temp' and 'windspeed' .And add lines for the Mean and Median on the plot. (Hint – Use matplotlib library hist to draw the plot. Use mean() and median() method to find mean & median)
3. Separate features and labels- Use X : features, y : labels.
$$X = ['season', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed']$$
$$y = ['rentals']$$
4. Split 70% data for training and 30% testing. (Use train_test_split in Python for split data).
5. Use Linear Regression to model the data set (Hint- Method LinearRegression())
6. Evaluate the trained models (Hint- Use model.predict() method for test data)
7. Use R-squared to find the accuracy of the model (Hint-r2_score() method to find the coefficient).

You need to submit the implemented code in Jupyter Notebook “.ipynb” file.