End to End Project - Bikes Assessment - Basic - Description

**Objective**

The objective of the project is - using historical usage patterns and weather data, forecast(predict) bike rental demand (number of bike users (‘cnt’)) on hourly basis.

Use the provided “Bikes Rental” data set to predict the bike demand (bike users count - 'cnt') using various best possible models (ML algorithms). Also, report the model that performs best, fine-tune the same model using one of the model fine-tuning techniques, and report the best possible combination of hyperparameters for the selected model. Lastly, use the selected model to make final predictions and compare the predicted values with the actual values.

Below are the details of the features list for the given Bikes data set:

1. instant: record index
2. dteday : date
3. season: season (1: springer, 2: summer, 3: fall, 4: winter)
4. yr: year (0: 2011, 1:2012)
5. mnth: month (1 to 12)
6. hr: hour (0 to 23)
7. holiday: whether the day is a holiday or not
8. weekday: day of the week
9. workingday: if day is neither weekend nor holiday is 1, otherwise is 0.
10. 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 + Thunderstorm + Mist, Snow + Fog
11. 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)
12. 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)
13. hum: Normalized humidity. The values are divided to 100 (max)
14. windspeed: Normalized wind speed. The values are divided to 67 (max)
15. casual: count of casual users
16. registered: count of registered users
17. cnt: count of total rental bikes including both casual and registered users

The "target" data set ('y') should have only one 'label' i.e. 'cnt'.  
  
**We will be following the below steps to solve this problem:**

1. Importing the libraries
2. Using some pre-defined utility functions
3. Loading the data
4. Cleaning the data
5. Dividing the dataset into training and test dataset
   * using train\_test\_split in the ratio 70:30
6. Training several models and analyzing their performance to select a model
7. Fine-tuning the model by finding the best hyper-parameters and features
8. Evaluating selected model using test dataset

**Acknowledgements**  
  
Cloudxlab is using this “Bike Sharing Demand” problem for its machine learning learners for learning and practicing. This dataset was provided by Hadi Fanaee Tork using data from [Capital Bikeshare](https://www.capitalbikeshare.com/system-data). We also thank the UCI machine learning repository for [hosting the dataset](http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset).  
Fanaee-T, Hadi, and Gama, Joao, Event labeling combining ensemble detectors and background knowledge, Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg.