Bike Rental Forecasting Project

This project is a simplified (basic) version of "Bike Rental" project (without any data analysis step).

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, and 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.

**Acknowledgements**:

This dataset was provided by Hadi Fanaee Tork using data from Capital Bikeshare. We also thank the UCI machine learning repository for hosting the 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.

**We will be following this example step-by-step in this order:**

1. Importing the libraries
2. Defining some utility functions
3. Loading the data
4. Cleaning the data
5. Adding derived features
6. Analyzing the dataset
7. Dividing the dataset into training and test dataset
8. Training several models and analyzing their performance
9. Selecting a model and evaluating using test dataset
10. Improving the model by finding the best hyper-parameters and features
11. Analyzing the residuals

**The dataset contains the following parameters:**

* 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)
* hr : hour (0 to 23)
* holiday : weather day is holiday or not (extracted from [Web Link])
* weekday : day of the week
* workingday : if day is neither weekend nor holiday is 1, otherwise is 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 Pallets + 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

**Added derived features and transforming the data**

These features are derived from the raw set of features:

1. **isWorking:** 1: Is a workingday and not a holiday, 0: Is not a workingday and is a holiday
2. **monthCount:** count of the number of months from the beginning of the dataset
3. **xformHr:** transform by shifting the hours by 5 hrs, if the hours are greater than 5, we subtract 5, else we add 19.
4. **dayCnt:** count of the days from the beginning of the dataset
5. **xformWorkHr:** transforming the hour dataset to make the non-working days to have hours from 25 to 48
6. **cntDeTrended:** De-trended count values

**ML algorithms used:**

1. Decision Tree Model
2. Linear Regression Model
3. Random Forest Model