Dear All,

Below is the description, checklist, responsibilities and guidelines for the Capstone Project. Please use this as the base of your Project Reports. Project reports has to be submitted before the finishing of the course or within 15 days of assigning:

Machine Learning Project Description :

1. Name of the Datasets : Household Power Consumption.
2. Dataset is a minute wise data logging of Power Consumption of a Household Consumption
3. Data Record Duration : December 2006 to November 2010 (4 Years).
4. Record Frequency : Per minute
5. Features recorded (07 No.) :
   1. Global Active Power
   2. Global Reactive Power
   3. Voltage
   4. Global Intensity
   5. Sub metering 1
   6. Sub metering 2
   7. Sub metering 3
6. Observations : 2075259 (around 2 million)

Objective of the project:

1. Develop a Machine Learning model to predict the Global Intensity
2. Develop a Machine Learning model to predict if the Global Intensity is greater than 15.

Project Activities and responsibilities :

1. Data loading
   1. Converting the original dataset from the raw txt file to csv format.
   2. Slice out the bottom/latest 200,000 observations for Project tests, and carry on the experiments with other available data observations.
   3. Crosscheck the discrepancies and abnormalities in the dataset, and amend for further usage.
   4. Load the data from csv file into Pandas dataframe for further usage.
2. Data Analysis
   1. In the prepared dataframe, checking the patterns of the different features using various reports and graphs
   2. Checking the correlations among features
   3. Checking the feature importance
   4. Checking the missing values
   5. Checking the outliers
   6. Statistical Analysis
3. Data Cleaning
   1. Dropping off the unnecessary or irrelevant features
   2. Treatment of the values creating the imbalance of the data
   3. Finding and treating the outliers wherever required
   4. Finding and treating the missing values appropriately wherever required.
4. Data Transformation
   1. Checking and Converting the features into needful formats.
   2. Merging the relevant features as per requirements
5. Pre-processing
   1. Prepare the data for modelling by using Train-Test split method
   2. Pre-processed data must contain X\_train, X\_test, Y\_train, Y\_test dataframes
6. Modelling
   1. Apply the relevant algorithms on the provided pre-processed data
   2. Evaluate the accuracy of the model by using test data
   3. Re-modelling to be done until the optimized results are achieved.
   4. Re-modelling to be done using Algorithms changes and Hyper-parameter finetuning methods.
7. Results
   1. Results of the machine learning to be drafted in the forms of Accuracy, Confusion Matrix, Classification reports wherever applied.