

Details of project requirements:

1. Identify the Input and Output Variables in your dataset.
2. If you have more than 10 input variables / features, select only up to 10 features which you think has relevance in your output variables. Or just randomly pick the 10 different features.
3. Perform further data cleaning in your dataset
 1. i. Identify input variables. List down the numeric features and non-numeric features. Then perform data transformation for non-numeric features. Such as ordinal or nominal values.
 2. ii. Identify and supply missing values in your input variables or features
 3. iii. If your output variables is in a form of text label, perform one-hot encoding.
4. Perform data standardization in your input variables and retain the value of your output variable. Then save to another csv files
5. Perform data normalization in your input variables retain the value of your output variable. Then save to another csv files
6. Use machine learning algorithm in your 3 datasets (raw, standardized data, normalized data)
 1. If your target variable is numeric, use the following learning algorithm
 1. Linear Regression (build, fit, transform)
 2. Lasso Regression (build, fit, transform)
 3. KNN algorithm (build, fit, transform)
 2. If your target variable is categorical, use the following learning algorithm
 1. Logistic Regression (build, fit, transform)
 2. Decision Trees (build, fit, transform)
 3. Random Forest algorithm (build, fit, transform)
7. Create a summary table for each dataset the performance of each learning Algorithm
 1. Predicting Numeric Values - Mean Squared Error for training and testing data
 2. Predicting Categorical Values - Accuracy for training and testing data

=====

Note:

1. Create a pdf file that captures the code and the corresponding output to each item stated above.
2. Files to be submitted online:
 1. Presentation file in a pdf or ppt form as a whole.
 2. Source Code (ipynb format)
 3. Dataset

- Raw
 - Standardized Dataset
 - Normalized dataset
- Submit the link of recorded video for each part. Assign each member for the video presentation, from data preprocessing up to 1 ML prediction for either standardized or normalized data.
 - Share your individual video to lovejhoye.raboy@ustp.edu.ph with the **subject: IT414-4R<your section> PROJECT - GRP# <Last Name>**. Example IT414-4R5 PROJECT GRP 1 RABOY
- Each member should allocate 2 minutes of presentation.

=====

Project Deadline - October 24, 2021 @11:59pm