CCPS 884
Data Mining Project
Andy Lee

TABLE OF CONTENTS

CCPS 844 Data Mining (Project) - Andy Lee

1. Select a dataset or datasets of your choice. Here are few links that can be helpful for you to select a dataset.
2. Once you have selected a dataset or datasets of your choice. After reading the datasets, check the type of different attributes/columns/features to ensure that you have appropriate types (categorical/numerical) for your columns.
3. Use visualization to understand your data
4. For exploratory analysis, apply clustering algorithms (K means/ Hierarchical clustering) to improve your understanding
5. Apply the concepts learned in Module 9 to select the features
6. Try to reduce the dimensions of the data if possible (Apply a dimensionality reduction algorithm). For step 7 use both the original data and the data that you get after applying the Step 6.
7. Divide your data in Train and Test or choose cross validation to evaluate the selected model
    - Apply all learned classification algorithms to choose which one performs best
    - Apply all learned regression algorithms to choose which one performs best

Please note that you need to get your data in appropriate format before applying a classification or regression algorithm. One of the differences is: class variable for a regression model is numeric whereas it is categorical for classification.