MACHINE LEARNING PROJECT PHASE 2

SUPERVISED/Unsupervised LEARNING ALGORITHMS

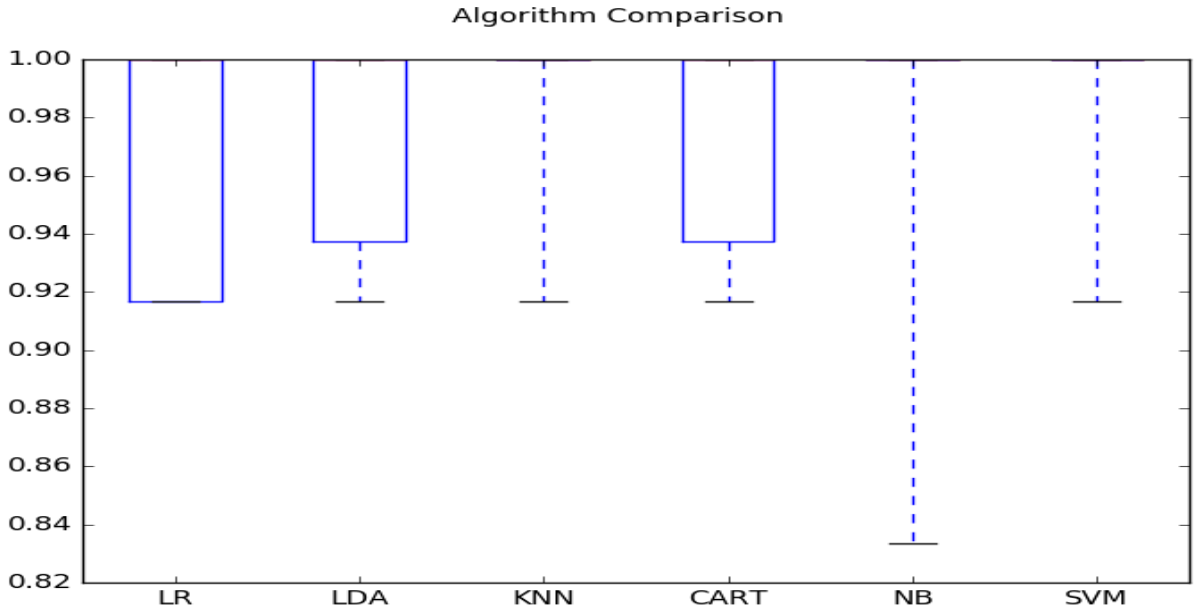
Submission Date: ~~30th November 2019~~

1. **Problem Definition**
    - **Give a concrete description of your machine learning problem in no more than 50 words**
2. **Datasets**
    - **Brief on the datasets used in the project.**
    - **At least 3 datasets should be chosen for the project.**
3. **Prepare Data**
    - **Explain the pre-processing done on your Dataset to make it suitable for applying machine learning algorithms.**
    - **Summarization:**
        - **Use statistical methods to understand the data and apply the required methods**
            - Dimensions of the dataset.
            - Statistical summary of all attributes.
            - Breakdown of the data by the class variable.
    - **Data Visualization:**
        - **Visualize the data using various plots like scatterplot, histograms, box plot etc and record your interpretations with varying values**
4. **Python packages**
    - **Brief on the python packages used for implementation of Machine learning algorithms pertaining to your project.**

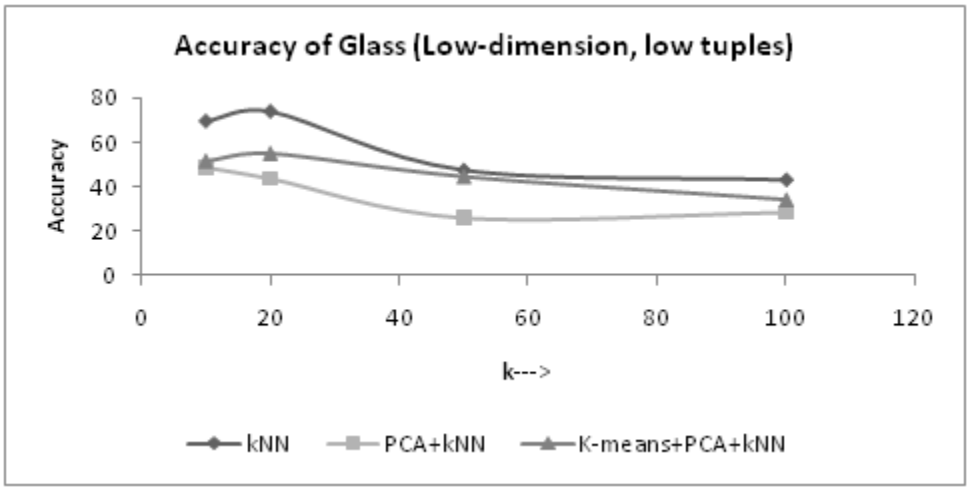
5. Supervised Learning Algorithms

○ At least 3 supervised machine learning algorithms are to be used

- Brief on each of supervised ML algorithms chosen for creating learning model from your dataset with proper justification
- Split your data into training, validation, and testing
- Use k-fold cross validation to evaluate your ML algorithm
- At least one supervised learning algorithm should be implemented.(Pl. choose the best possible supervised learning algorithm for your project)
- Create models of the data and estimate their accuracy on unseen data using the specified ML algorithms.
    - Example: If Logistic regression, SVM, and kNN are used for classification, create models for different algorithms. Select the best model.
    - Plot a comparison graph showing the accuracy comparison of various algorithms on each of your datasets.
      Example : Following figure shows an instance of average accuracy comparison of various algorithms on a particular dataset.



- Make Predictions on validation dataset. Plot accuracy and time for varying parameters
    - Example : If kNN is used, parameters to be used are various values of k. Plot accuracy for different values of k. See below figure for a sample. Similarly plot execution time for different values of k.



6. Unsupervised Learning Algorithms

- **At least 2 unsupervised algorithms should be used**
    - **Brief on each of the unsupervised ML algorithms used to analyze dataset with proper justification.**
    - **Evaluate the analyzed dataset with proper metrics**