Data Mining and Machine Learning
Assignment #4

1. Find the following data files in the shared folder /data on OneDrive
   - credit-g.arff, Diabetes.arff, breast-cancer.arff, vote.arff, and labor.arff

   The data files can be found in Weka's installation folder.

2. Read the descriptions of the dataset and attributes in each data file, and understand the datasets.
3. Work on one of the five data sets.
   - Use three different types of classification methods discussed in this course.

   The following classifiers in scikit-learn could be used.
   - Wittgenstein (Rule-based)
   - DecisionTreeClassifier (Decision Tree)
   - CategoricalNB (Bayes)
   - LogisticRegression (Functions)
   - MLPClassifier (Neural Network)

   The following classifiers in Weka could be used.
   - JRip (Rules)
   - J48 (Decision Tree)
   - NaïveBayes (Bayes)
   - Logistic (Functions)
   - MultilayerPerceptron (Neural Network)
   - Evaluate the performance of the classifiers

Report (50 points)

The report consists of description of dataset and task, justification for your choice of classification methods, data preprocessing, performance evaluation design, and results.

Bonus (5 points for final grade)

- Complete this project using Python.

Turn in your report on Canvas.

Let me know ASAP if you would like to work on other datasets. And you have to show me the dataset you would like to work on.

Also, let me know ASAP if you would like to team up for working on big or challenging projects want to team up