Homework #3 – ML Classifiers

Note:
Submission of homework must be electronic. Problems require you to write computer program. Please submit your code and a PDF file containing output of your code or a Jupyter notebook that contains the output.

Experiments with different ML classifiers:
Use the Concrete Quality dataset given in Concrete_Quality_Binary.xlsx.

Note: The file is of 'xlsx' format. It has 639 labeled samples with 9 features – Index, Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, Age. The last column represents the quality of cement (Good/Bad).

- Scale the data using standard scaler and implement 7 models listed below.
- Use K-fold cross validation strategy with K=5 and find the best classifier for this dataset. Make sure to shuffle the data while doing cross-validation.
- Try to optimize the algorithms by tuning the hyper-parameters (e.g. k for KNN, max_depth for Decision Tree etc.) for the classifiers.
  1) Logistic Regression
  2) Naive Bayes
  3) Decision Tree
  4) Random Forest
  5) K-Nearest Neighbor
  6) Linear SVM
  7) GridSearchCV with Random Forest

Submission:
1. Report the average accuracies across 5 runs for every model.
2. For the best model, report the confusion matrix and its accuracy.
3. Code with output (A Jupyter notebook with output).