ECE657A Assignment 2
Submitted by :
Amandeep Kaur (21044104)
Bhupesh Dod (21046099)

Question 6: Final Results

Solution:

The final results are shown in the following tables.

Table 6.1 Accuracy Percentage table for different algorithms with best settings on Abalone Dataset

	Best Settings			Abalone_Raw	Abalone_PCA	Abalone_LDA
KNN	k=73, weights = "distance"			28.32%	27.17%	27.84%
Complement Naive Bayes	fit_prior=True			17.56%	20.02%	23.53%
Decision Tree	RAW max_depth = 5	PCA max_depth = 4	LDA max_depth = 4	24.69%	27.94%	24.69%
Random Forest	RAW max_depth = 9, n_estimators = 50	PCA max_depth = 4, n_estimators = 40	LDA max_depth = 6, n_estimators = 40	27.18%	26.79%	27.94%
Gradient Tree Boosting	RAW max_depth = 5, max_leaf_nodes= 4, n_estimators = 7	PCA max_depth = 4, max_leaf_nodes = 4, n_estimators = 5	LDA max_depth = 2, max_leaf_nodes = 16, n_estimators = 10	26.60%	23.72%	26.70%

Table 6.2 Accuracy Percentage table for different algorithms with best settings on Wine Dataset

	Best Settings			Wine_Raw	Wine_PCA	Wine_LDA
KNN	k=46 weights = "distance"			68.15%	67.38%	68.12%
MultiNomial Naive Bayes	fit_prior=True			47.89%	44.46%	44.46%
Decision Tree	RAW max_depth = 13	PCA max_depth = 14	LDA max_depth = 17	58.40%	58.65%	59.38%
Random Forest	RAW max_depth = 14, n_estimators = 50	PCA max_depth = 14, n_estimators = 50	LDA max_depth = 14, n_estimators = 40	68.92%	67.57%	64.49%
Gradient Tree Boosting	RAW max_depth = 6, max_leaf_nodes= 32, n_estimators = 10	PCA max_depth = 7, max_leaf_nodes = 32, n_estimators = 10	LDA max_depth = 8, max_leaf_nodes = 32, n_estimators = 10	59.57%	56.55%	57.97%

It is evident from the above findings that for both the datasets, maximum accuracy is given by KNN Algorithm when applied on abalone data. The next best accuracy is of Random Forest. The algorithm to produce least accuracy is Naive Bayes be it Complement Naive Bayes or Multinomial Naive Bayes for both the datasets. Out of all the Decision Tree and Decision Tree Ensembles Random Forest has performed the best. Random Forest has emerged to better in performance and in runtime than Gradient Tree Boosting. Random forest has performed the best on Wine data overall. Amongst KNN and Naive Bayes, KNN has performed better on these particular datasets, which indicates that Naive Bayes is not a good choice of algorithm for categorical data.

Best Algorithm for Abalone : KNN
Worst Algorithm for Abalone : Naive Bayes

Best Algorithm for Wine : Random Forest
Worst Algorithm for Wine : Naive Bayes