CS412 Assignemnt4 report

1. brief introduction of the classification methods in your classification framework

(1) Basic Classification Method: I create a class called DecisionTreeClassifier including fit, predict and build_tree functions and a node class. Then I write other necessary functions such as load data and get gini. First, I create a DecisionTreeClassifier and fit the train data by calling fit function. After fitting data, it will generate a decision tree. Then, I load test data and call predict function to test. Besides, I set depth of the decision tree 20.

(2) Ensemble Classification Method: I import DecisionTreeClassifier from decision tree file and create a class called RandomForestClassifier including fit, predict and score functions. Also, I create other necessary functions such as load data. First, I create a RandomForestClassifier and fit the train data by calling fit function. After fitting data, it will generate a decision tree. Then, I load test data and call predict function to test. Besides, I set bootstrap 0.632.

2. all model evaluation measures you calculated above (7 metrics * 2 methods * 4 datasets)

Train:

DT	Sensitivity	Specificity	Precision	Recall	F-1	F-0.5	F-2	Accuracy
Balance-1	1	1	1	1	1	1	1	1
Balance-2	1	1	1	1	1	1	1	
Balance-3	1	1	1	1	1	1	1	
Nursery-1	1	1	1	1	1	1	1	1
Nursery-2	1	1	1	1	1	1	1	
Nursery-3	1	1	1	1	1	1	1	
Nursery-4	1	1	1	1	1	1	1	
Nursery-5	1	1	1	1	1	1	1	
Led-1	0.777	0.895	0.767	0.777	0.772	0.679	0.894	0.858
Led-2	0.895	0.777	0.901	0.895	0.898	0.846	0.956	
Poker-1	1	1	1	1	1	1	1	1
Poker-2	1	1	1	1	1	1	1	1

RF	Sensitivity	Specificity	Precision	Recall	F-1	F-0.5	F-2	Accuracy
Balance-1	1	1	1	1	1	1	1	1
Balance-2	1	1	1	1	1	1	1	
Balance-3	1	1	1	1	1	1	1	
Nursery-1	1	1	1	1	1	1	1	1
Nursery-2	1	1	1	1	1	1	1	
Nursery-3	1	1	1	1	1	1	1	
Nursery-4	1	1	1	1	1	1	1	
Nursery-5	1	1	1	1	1	1	1	
Led-1	0.777	0.895	0.767	0.777	0.772	0.679	0.894	0.858
Led-2	0.895	0.777	0.901	0.895	0.898	0.846	0.956	
Poker-1	1	1	1	1	1	1	1	1
Poker-2	1	1	1	1	1	1	1	

Test:

DT	Sensitivity	Specificity	Precision	Recall	F-1	F-0.5	F-2	Accuracy
Balance-1	0.831	0.734	0.848	0.831	0.771	0.677	0.893	0.777
Balance-2	0.892	0.682	0.812	0.892	0.784	0.694	0.901	
Balance-3	0	0.862	0	0	0	0	0	
Nursery-1	0.960	0.982	0.963	0.960	0.963	0.954	0.968	0.974
Nursery-2	1	1	1	1	1	1	1	
Nursery-3	0.971	0.977	0.967	0.971	0.961	0.939	0.972	
Nursery-4	0.892	0.977	0.901	0.892	0.755	0.581	0.826	
Nursery-5	0	0.974	0	0	0	0	0	
Led-1	0.781	0.893	0.765	0.781	0.773	0.680	0.895	0.858
Led-2	0.893	0.781	0.901	0.893	0.897	0.844	0.956	
Poker-1	0.315	0.686	0.324	0.315	0.319	0.280	0.540	0.566
Poker-2	0.686	0.315	0.677	0.686	0.682	0.576	0.843	

RF	Sensitivity	Specificity	Precision	Recall	F-1	F-0.5	F-2	Accuracy
Balance-1	0.811	0.741	0.882	0.811	0.762	0.668	0.890	0.773
Balance-2	0.901	0.667	0.793	0.901	0.793	0.705	0.889	
Balance-3	0	0.857	0	0	0	0	0	
Nursery-1	0.955	0.980	0.960	0.955	0.967	0.958	0.972	0.972
Nursery-2	1	1	1	1	1	1	1	
Nursery-3	0.966	0.976	0.962	0.966	0.958	0.950	0.964	
Nursery-4	0.908	0.973	0.906	0.908	0.634	0.520	0.813	
Nursery-5	0	0.972	0	0	0	0	0	
Led-1	0.781	0.893	0.765	0.781	0.773	0.680	0.895	0.858
Led-2	0.893	0.781	0.901	0.893	0.897	0.844	0.927	
Poker-1	0.854	0.255	0.706	0.854	0.773	0.731	0.819	0.660
Poker-2	0.255	0.854	0.455	0.225	0.327	0.393	0.280	

3. parameters you chose during implementation and why you chose these parameters

After testing multi-times, I find 20 is appropriate number for max depth. And I set bootstrap 0.632 since it is recommended on the slide and it works well.

4. and also your conclusion on whether the ensemble method improves the performance of the basic classification method you chose, why or why not

Yes, I think ensemble method improves the performance of the basic classification. As we can see, for poker data set, it improves a lot.