ACCURACY REPORT FOR ALL DATASET

DATASET – 1

METHOD	BEST PARAM.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	% AVG ACCURACY
Decision Tree	Cp=0.03	100	100	99.5	100	100	100	100	100	99.5	99.5	99.5
SVM	Kernel=default	99	98.5	96.5	98	99	97.5	99	99	98	98.5	98.3
Naïve Bayesian	Laplace=5	98	96	96.5	97	97	96.5	97.5	96.5	96.5	99	97.05
Knn	K=11	80	85.5	80	83	84	84.5	79.5	84.5	85	85.5	83.25
Logistic Regression	threshold=0.90	79.55	80.38	79.77	80.22	79.55	56	80.22	81.5	82.77	81.88	93.7
Neural Network	Hidden=6	67.5	62.5	62.5	70	67.5	57.5	60	72.5	67.5	67.5	85.75
Bagging	Maxdepth=1	99.981	99.994	99.950	99.45	99.45	99.84	99.96	99.87	99.99	99.98	99.96
Random Forest	Ctree_control= default	100	100	100	100	99.96	100	99.97	100	99.96	100	99.993
Boosting	Iter=20	100	100	99.5	100	100	100	100	100	99.5	100	99.95

DATASET – 2

METHOD	BEST PARAM.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	AVERAGE ACCURACY
Decision Tree	Cp=0.03	70	62.5	57.5	75	67.5	72.5	60	80	67.5	75	68.75
SVM	Kernel=default	70	67.5	62.5	82.5	67.5	62.5	60	75	70	70	68.75
Naïve Bayesian	Laplace=5	65	67.5	62.5	75	60	72.5	60	70	67.5	70	67
Knn	K=5	90	82.5	82.5	87.5	87.5	87.5	85	87.5	92.5	87.5	87
Logistic Regression	threshold=0.65	66.388	68.611	70.555	66.11	67.5	69.444	68.611	67.77	68.333	65	67.83
Neural Network	Hidden=4	67.5	62.5	62.5	70	67.5	57.5	60	72.5	67.5	67.5	65.5
Bagging	Maxdepth=1	70.30	69.33	67.07	69.15	68.9	71.62	67.84	82.32	0	80.84	64.74
Random Forest	Ctree_control=default	82.17	82.17	83.45	81.25	86.7	84.80	89.1	81.25	83.45	83.45	84.90
Boosting	Iter=20	72.5	57.5	70	65	67.5	65	55	67.5	75	65	66

DATASET-3

METHOD	BEST PARAM.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	AVERAGE ACCURACY
Decision Tree	Cp=0.03	95	70	65	80	85	55	65	65	75	82	73.5
SVM	Kernel=default	70	80	80	80	80	70	75	60	70	65	73
Naïve Bayesian	Laplace=5	60	75	70	75	75	60	60	55	55	75	66
Knn	K=5	75	60	75	65	80	75	65	65	80	70	71
Logistic Regression	threshold=0.40	75	65	75	65	65	65	55	65	65	70	64.5
Neural Network	Hidden=10	70	80	80	80	80	70	70	60	70	65	72.5
Bagging	Maxdepth=1	55	75	85	65	75	65	65	55	65	60	66.5
Random Forest	Ctree_control=default	75	80	80	80	75	70	55	60	70	65	71
Boosting	Iter=20	65	75	75	80	80	85	65	80	75	75	75.5

DATASET – 4

METHOD	BEST PARAM.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	AVERAGE ACCURACY
Decision Tree	Cp=0.03	94.73	91.22	94.73	96.49	89.47	98.24	96.49	96.49	94.73	92.22	94.91
SVM	Kernel=default	100	98.2	98.24	100	98.24	98.24	98.24	96.49	100	96.49	98.42
Naïve Bayesian	Laplace=5	98.24	92.98	92.98	96.49	91.22	94.73	98.24	89.47	96.49	94.73	94.56
Knn	K=5	71.92	78.94	675.43	84.21	73.68	78.9	75.43	82.45	85.96	77.19	78.42
Logistic Regression	threshold=0.60	94.73	92.98	96.49	96.49	98.24	94.73	98.24	98.24	96.49	96.49	96.3
Neural Network	Hidden=10	63.15	64.91	66.66	57.89	61.40	64.91	66.66	61.40	66.66	61.40	63.50
Bagging	Maxdepth=1	95.2	93.73	92.1	93.27	95.67	90.32	90.32	95.2	93.73	95.2	93.68
Random Forest	Ctree_control=default	92.98	91.22	94.73	91.22	96.49	92.98	100	96.73	94.73	98.24	94.38
Boosting	Iter=20	96.49	96.49	96.49	94.73	94.73	96.49	100	94.73	94.73	98.24	96.31

DATASET – 5

METHOD	BEST PARAM.	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	AVERAGE ACCURACY
Decision Tree	Cp=0.03	91.66	97.22	88.88	88.88	88.88	89	80.55	83.33	88.88	91.66	90
SVM	Kernel=default	91.22	93.22	86.11	94.44	86.11	88.88	91.66	97.22	96.78	92.22	90.61
Naïve Bayesian	Laplace=5	94.44	86.11	83.33	91.66	86.11	88.88	91.66	94.44	86.11	94.44	89.72
Knn	K=9	92.56	83.33	91.66	83.33	91.66	91.66	80.55	94.44	94.44	88.88	89.44
Logistic Regression	threshold=0.85	72.22	72.22	69.44	61.11	69.44	61.11	72.22	69.44	80.55	69.44	69.72
Neural Network	Hidden=8	41.66	44.44	33.33	41.66	38.88	27.77	41.66	27.77	41.66	47.22	38.61
Bagging	Maxdepth=1	94.44	91.66	94.44	91.66	86.11	88.88	86.11	88.8	91.66	86.11	90.55
Random Forest	Ctree_control=default	88.88	91.66	83.33	88.88	89	86.11	88.88	83.33	97.22	91.66	89.44
Boosting	Iter=20	94.44	94.44	88.88	88.88	80.55	86.11	88.88	100	94.44	94.44	91.11

OBSERVATIONS FROM ACCURACY TABLES:

- DATASET - 1

For dataset 1, all classifiers worked equally good. Decision tree, Bagging, Random Forest, Boosting had nearly perfect accuracy of around 99% for this dataset. Other classifiers such as SVM, Naive Bayesian also had accuracy of around 98%.

Best classifier - DT, Bagging, RF, Bosting

- DATATSET - 2

For dataset - 2, k-nearest neighbour had the best accuracy of 87%, followed closely by Random Forest giving accuracy of 85%. Whereas ensemble methods like bagging and boosting did not perform well on this dataset.

Best classifier - k-nearest neighbour

- DATATSET - 3

For dataset-3, all classifiers worked averagely with no higher accuracy reported.

Best classifier - Boosting

- DATATSET - 4

For dataset-4, SVM reported the highest accuracy of 98% followed by boosting. All ensemble methods performed well for this dataset.

Best Classifier - SVM

- DATATSET - 5

For dataset - 5, boosting and svm showed comparable accuracy. Neural network demonstrated the lowest accuracy for this dataset with hidden layers as 8. With increase/decrease of hidden layers, accuracy was decreasing. Best accuracy for NN was reported at hidden layers =8 which was 38%.

Best classifier - Boosting