Individual Project I: Classification - Report

Ripper

Dataset: IRIS

Prediction Table and Measures

Predictions	setosa	versicolor	viginica
setosa	9	1	0
versicolor	0	9	2
virginica	0	0	9

	Accuracy	Precision	Recall	F Measure
setosa	0.9666667	0.9000000	1.0000000	0.9473684
versicolor	0.9000000	0.8181818	0.9000000	0.8571429
virginica	0.9333333	1.0000000	0.8181818	0.9000000

Confusion Matrices

Predicted	setosa	not setosa
setosa	9	0
not setosa	1	20

Predicted	versicolor	not versicolor
versicolor	9	1
not versicolor	2	18

Predicted	virginica	not virginica
virginica	9	2
not virginica	0	19

Total accuracy of Ripper for the test set: 0.9

Dataset: Life Expectancy

Prediction Table and Measures

Predictions	Africa	Asia	Europe	N.A	S.A
Africa	7	5	3	4	0
Asia	1	8	0	0	0
Europe	0	4	6	0	1
N.A	0	0	0	0	0
S.A	0	0	0	0	0

	Accuracy	Precision	Recall	F Measure
Africa	0.6666667	0.3684211	0.8750000	0.5185185
Asia	0.7435897	0.8888889	0.4705882	0.6153846
Europe	0.7948718	0.5454545	0.6666667	0.6000000
N.A	0.8974359	NaN	0.0000000	NaN
S.A	0.9743590	NaN	0.0000000	NaN

Confusion Matrices

Predicted	Africa	not Africa
Africa	7	1
not Africa	12	19

Predicted	Asia	not Asia
Asia	8	9
not Asia	1	21

Predicted	Europe	not Europe
Europe	6	3
not Europe	5	25

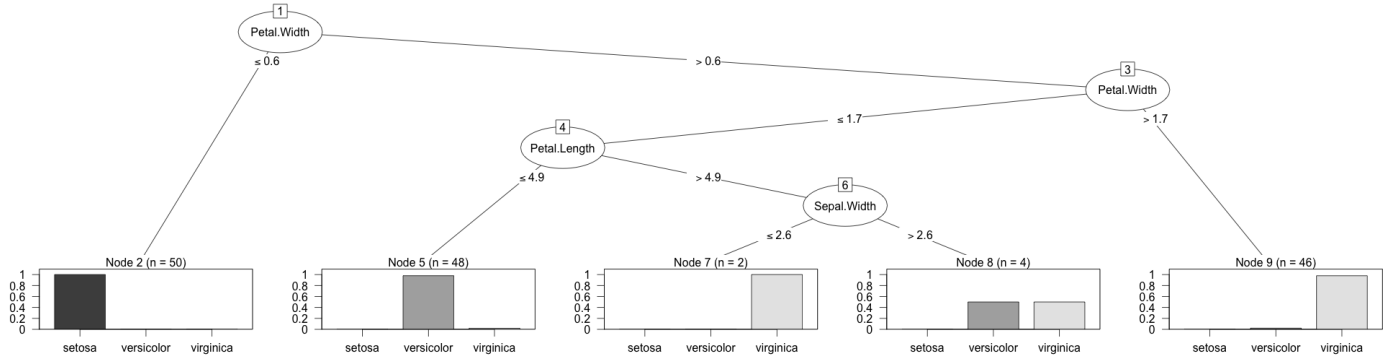
Predicted	N.A	not N.A
N.A	0	4
not N.A	0	35

Predicted	S.A	not S.A
S.A	0	1
not S.A	0	38

Total accuracy of Ripper for the test set: 0.538461538461538

C45

Dataset: IRIS



Prediction Table and Measures

Predictions	setosa	versicolor	viginica
setosa	9	0	0
versicolor	0	10	2
virginica	0	0	9

	Accuracy	Precision	Recall	F Measure
setosa	1.0000000	1.0000000	1.0000000	1.0000000
versicolor	0.9333333	0.8333333	1.0000000	0.9090909
virginica	0.9333333	1.0000000	0.8181818	0.9000000

Confusion Matrices

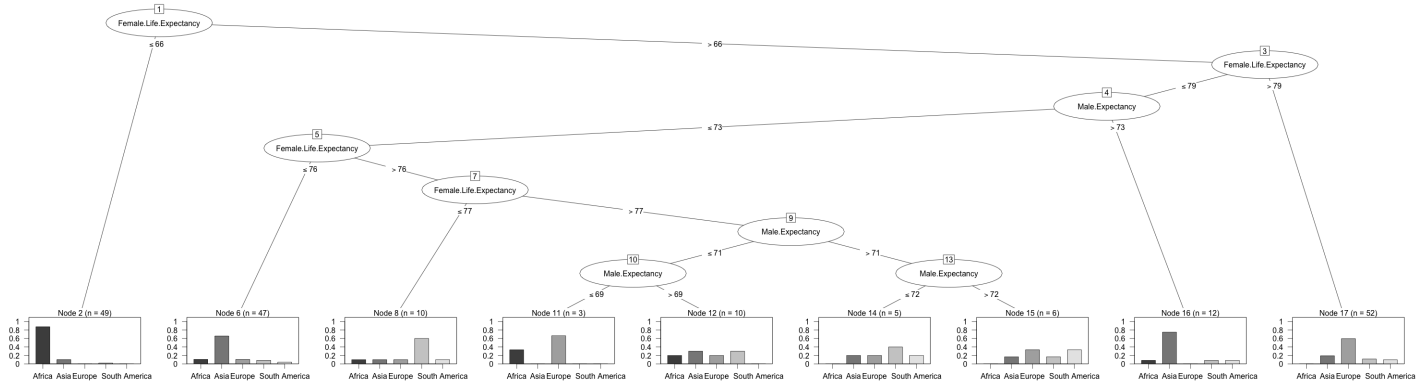
Predicted	setosa	not setosa
setosa	9	0
not setosa	0	21

Predicted	versicolor	not versicolor
versicolor	10	0
not versicolor	2	18

Predicted	virginica	not virginica
virginica	9	2
not virginica	0	19

Total accuracy of C45 for the test set: 0.933333333333333

Dataset: Life Expectancy



Prediction Table and Measures

Predictions	Africa	Asia	Europe	N.A	S.A
Africa	7	2	1	1	0
Asia	1	9	1	1	0
Europe	0	4	6	0	1
N.A	0	2	0	1	0
S.A	0	0	1	1	0

	Accuracy	Precision	Recall	F Measure
Africa	0.8717949	0.6363636	0.8750000	0.7368421
Asia	0.7179487	0.7500000	0.5294118	0.6206897
Europe	0.7948718	0.5454545	0.6666667	0.6000000
N.A	0.8717949	0.3333333	0.2500000	0.2857143
S.A	0.9230769	0.0000000	0.0000000	NaN

Confusion Matrices

Predicted	Africa	not Africa
Africa	7	1
not Africa	4	27

Predicted	Asia	not Asia
Asia	9	8
not Asia	3	19

Predicted	Europe	not Europe
Europe	6	3
not Europe	5	25

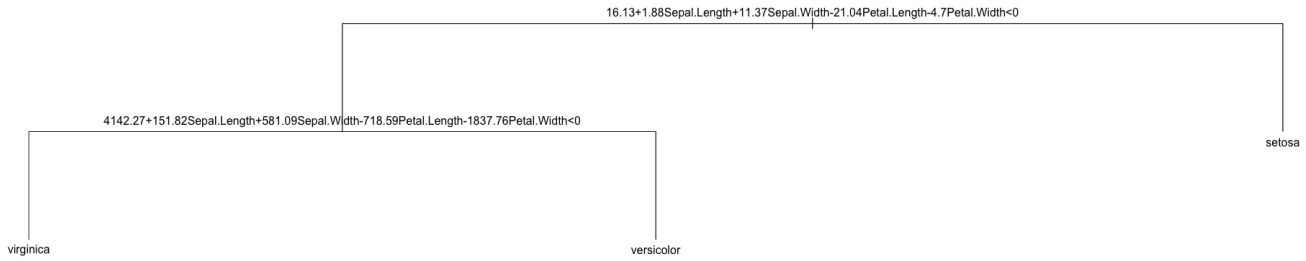
Predicted	N.A	not N.A
N.A	1	3
not N.A	2	33

Predicted	S.A	not S.A
S.A	0	1
not S.A	2	36

Total accuracy of C45 for the test set: 0.58974358974359

Oblique Tree

Dataset: IRIS



Prediction Table and Measures

Predictions	setosa	versicolor	viginica
setosa	9	0	0
versicolor	0	10	2
virginica	0	0	9

	Accuracy	Precision	Recall	F Measure
setosa	1.0000000	1.0000000	1.0000000	1.0000000
versicolor	0.9333333	0.8333333	1.0000000	0.9090909
virginica	0.9333333	1.0000000	0.8181818	0.9000000

Confusion Matrices

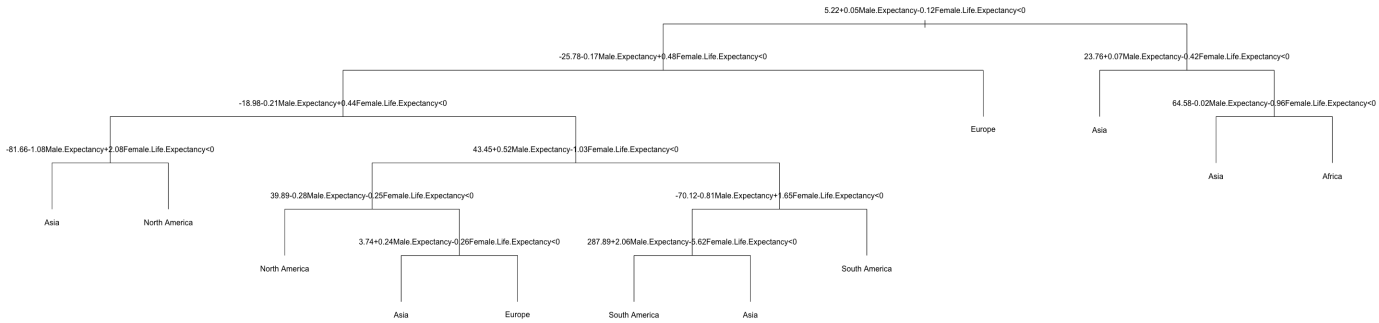
Predicted	setosa	not setosa
setosa	9	0
not setosa	0	21

Predicted	versicolor	not versicolor
versicolor	10	0
not versicolor	2	18

Predicted	virginica	not virginica
virginica	9	2
not virginica	0	19

Total accuracy of Oblique for the test set: 0.933333333333333

Dataset: Life Expectancy



Prediction Table and Measures

Predictions	Africa	Asia	Europe	N.A	S.A
Africa	6	2	0	1	0
Asia	2	9	2	0	0
Europe	0	4	7	0	1
N.A	0	0	0	2	0
S.A	0	2	0	1	0

	Accuracy	Precision	Recall	F Measure
Africa	0.8717949	0.6666667	0.7500000	0.7058824
Asia	0.6923077	0.6923077	0.5294118	0.6000000
Europe	0.8205128	0.5833333	0.7777778	0.6666667
N.A	0.9487179	1.0000000	0.5000000	0.6666667
S.A	0.8974359	0.0000000	0.0000000	NaN

Confusion Matrices

Predicted	Africa	not Africa
Africa	6	2
not Africa	3	28

Predicted	Asia	not Asia
Asia	9	8
not Asia	4	18

Predicted	Europe	not Europe
Europe	7	2
not Europe	5	25

Predicted	N.A	not N.A
N.A	2	2
not N.A	0	35

Predicted	S.A	not S.A
S.A	0	1
not S.A	3	35

Total accuracy of Oblique for the test set: 0.615384615384615

Naive Bayes

Dataset: IRIS

Prediction Table and Measures

Predictions	setosa	versicolor	viginica
setosa	9	0	0
versicolor	0	10	1
virginica	0	0	10

	Accuracy	Precision	Recall	F Measure
setosa	1.0000000	1.0000000	1.0000000	1.0000000
versicolor	0.9666667	0.9090909	1.0000000	0.952381
virginica	0.9666667	1.0000000	0.9090909	0.952381

Confusion Matrices

Predicted	setosa	not setosa
setosa	9	0
not setosa	0	21

Predicted	versicolor	not versicolor
versicolor	10	0
not versicolor	1	19

Predicted	virginica	not virginica
virginica	10	1
not virginica	0	19

Total accuracy of NaiveBayes for the test set: 0.966666666666667

Dataset: Life Expectancy

Prediction Table and Measures

Predictions	Africa	Asia	Europe	N.A	S.A
Africa	7	4	0	1	0
Asia	0	7	2	0	0
Europe	0	4	6	0	1
N.A	1	2	1	3	0
S.A	0	0	0	0	0

	Accuracy	Precision	Recall	F Measure
Africa	0.8461538	0.5833333	0.8750000	0.7000000
Asia	0.6923077	0.7777778	0.4117647	0.5384615
Europe	0.7948718	0.5454545	0.6666667	0.6000000
N.A	0.8717949	0.4285714	0.7500000	0.5454545
S.A	0.9743590	NaN	0.0000000	NaN

Confusion Matrices

Predicted	Africa	not Africa
Africa	7	1
not Africa	5	26

Predicted	Asia	not Asia
Asia	7	10
not Asia	2	20

Predicted	Europe	not Europe
Europe	6	3
not Europe	5	25

Predicted	N.A	not N.A
N.A	3	1
not N.A	4	31

Predicted	S.A	not S.A
S.A	0	1
not S.A	0	38

Total accuracy of NaiveBayes for the test set: 0.58974358974359

kNN

Dataset: IRIS

Prediction Table and Measures

Predictions	setosa	versicolor	viginica
setosa	9	0	0
versicolor	0	10	1
virginica	0	0	10

	Accuracy	Precision	Recall	F Measure
setosa	1.0000000	1.0000000	1.0000000	1.0000000
versicolor	0.9666667	0.9090909	1.0000000	0.952381
virginica	0.9666667	1.0000000	0.9090909	0.952381

Confusion Matrices

Predicted	setosa	not setosa
setosa	9	0
not setosa	0	21

Predicted	versicolor	not versicolor
versicolor	10	0
not versicolor	1	19

Predicted	virginica	not virginica
virginica	10	1
not virginica	0	19

Total accuracy of KNN for the test set: 0.966666666666667

Dataset: Life Expectancy

Prediction Table and Measures

Predictions	Africa	Asia	Europe	N.A	S.A
Africa	7	2	0	1	0
Asia	0	9	1	1	0
Europe	0	5	7	0	1
N.A	1	1	1	1	0
S.A	0	0	0	1	0

	Accuracy	Precision	Recall	F Measure
Africa	0.8974359	0.7000000	0.8750000	0.7777778
Asia	0.7435897	0.8181818	0.5294118	0.6428571
Europe	0.7948718	0.5384615	0.7777778	0.6363636
N.A	0.8461538	0.2500000	0.2500000	0.2500000
S.A	0.9487179	0.0000000	0.0000000	NaN

Confusion Matrices

Predicted	Africa	not Africa
Africa	7	1
not Africa	3	28

Predicted	Asia	not Asia
Asia	9	8
not Asia	2	20

Predicted	Europe	not Europe
Europe	7	2
not Europe	6	24

Predicted	N.A	not N.A
N.A	1	3
not N.A	3	32

Predicted	S.A	not S.A
S.A	0	1
not S.A	1	37

Total accuracy of KNN for the test set: 0.615384615384615

Comparison of various algorithms

Accuracy	IRIS	Life Expectancy
Ripper	0.9	0.538461538461538
C45	0.933333333333333	0.58974358974359
Oblique	0.933333333333333	0.615384615384615
Naive Bayes	0.966666666666667	0.58974358974359
kNN	0.966666666666667	0.615384615384615

Conclusion

From the above comparison table, we can observe that the accuracy for a given data set is nearly same for various classification algorithms. Hence, we can conclude that data plays a more important role in classification, as compared to the classification algorithm.