Report Individual Project

Name: Rohit Kumar Malik
UFID: 94711368

Steps to run the script:

1. install the following packages.

caret, datasets, rpart, klaR, RWeka, MASS, e1071, ggplot2, tree, party, oblique.tree, partykit

2. put the life_expectancy.csv file in along with the script file.
3. to install all the packages please uncomment all the install.packages statement
4. launch and run the project1.R script to produce the results.

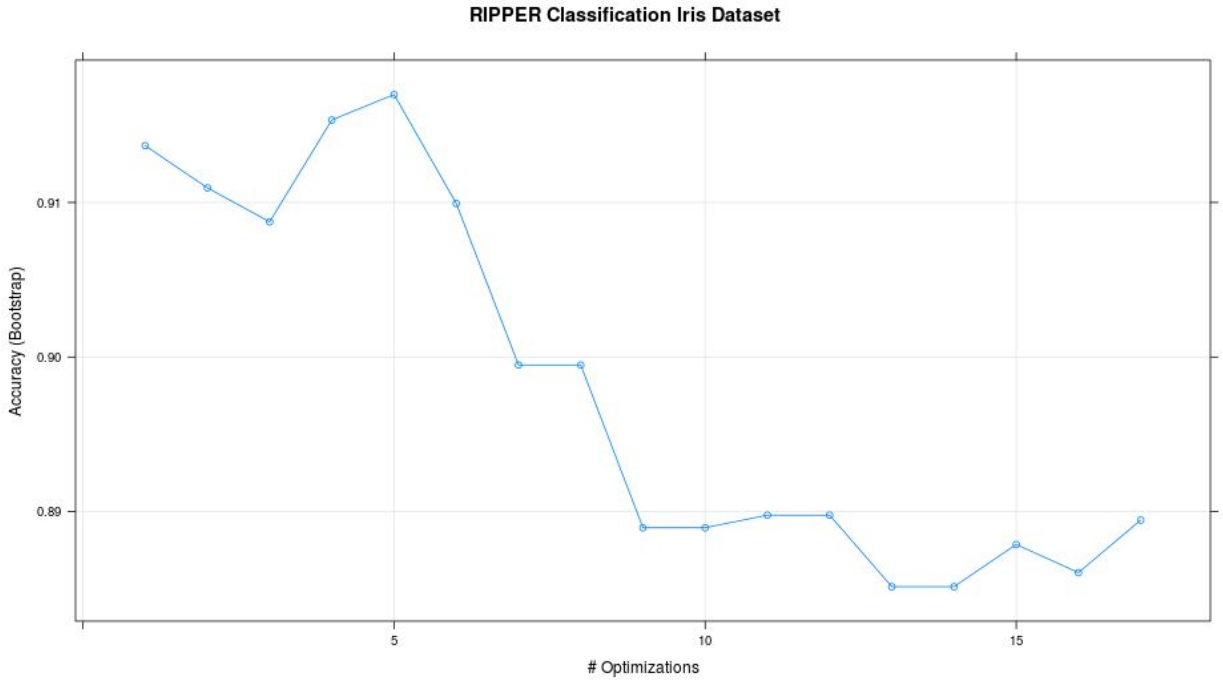
Note: I have removed the functions as I was facing issues in reproducing the results. All the classifications are separated by comments

Results:

Iris Dataset Analysis:

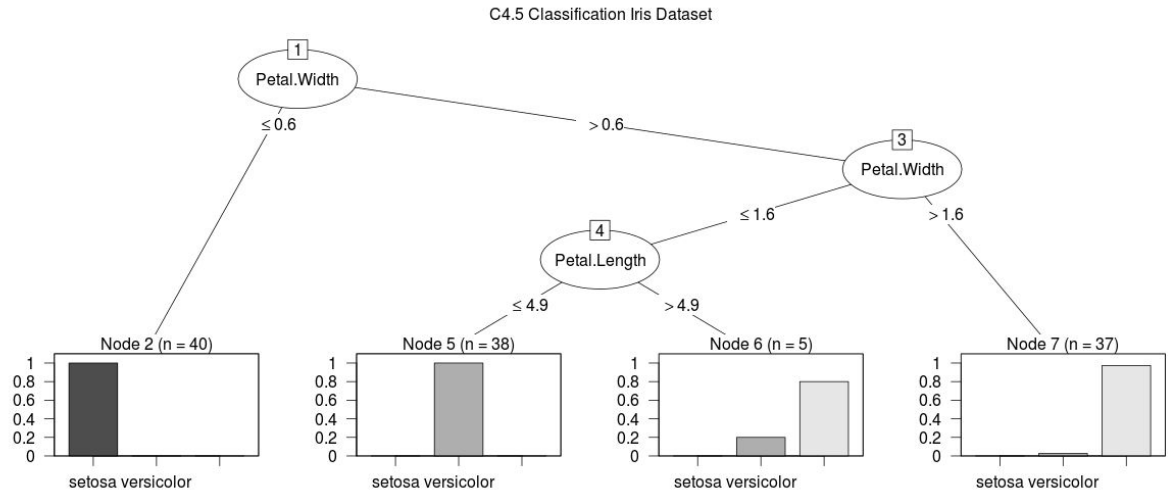
Desion Tree: RIPPER

Prediction	**setosa**	**versicolor**	**virginica**
setosa	10	0	0
versicolor	0	8	0
virginica	0	2	10
Accuracy :	0.9333		
Recall	1.0000	0.8000	1.0000
Precision	1.0000	1.0000	0.8333
F-Measure	1.0000	.89	.907



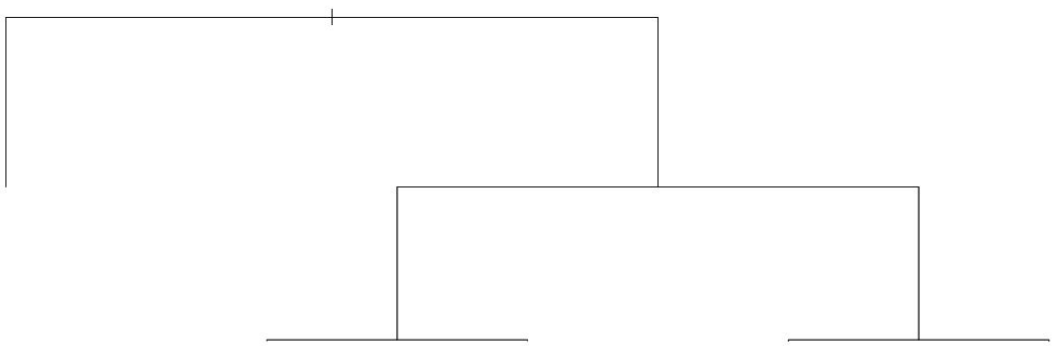
C4.5: Confusion Matrix

Prediction	**setosa**	**versicolor**	**virginica**
setosa	10	0	0
versicolor	0	9	0
virginica	0	1	10
Accuracy :	0.9667		
Recall:	1.0000	0.9000	1.0000
Precision:	1.0000	1.0000	0.9091
F-Measure:	1.0000	.95	.952



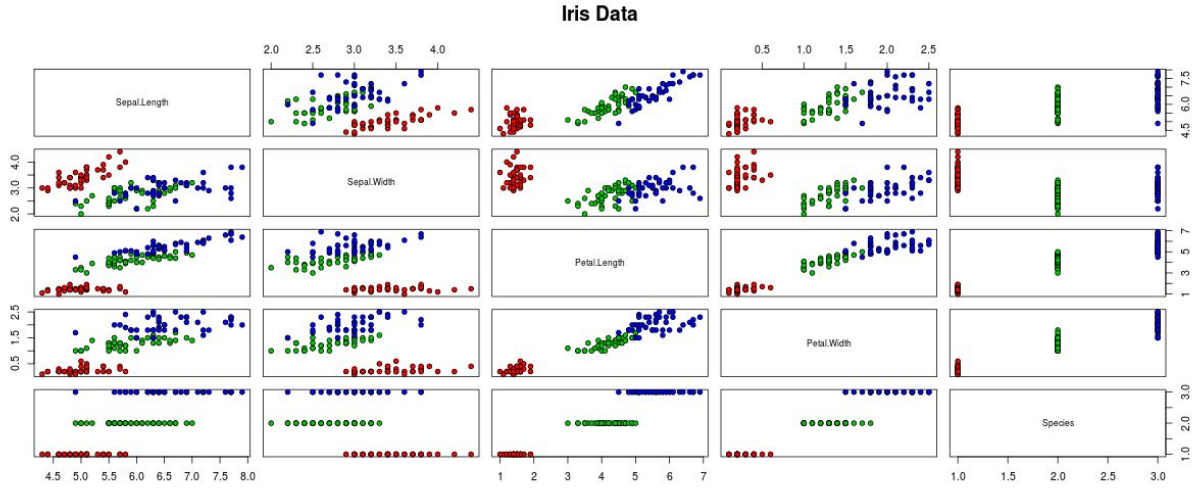
ObliqueTree:Confusion Matrix

	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	9	0
virginica	0	1	10
Accuracy :	0.9667		
Recall:	1.0000	0.9000	1.0000
Precision:	1.0000	1.0000	0.9091
F-Measure:	1.0000	.95	.952



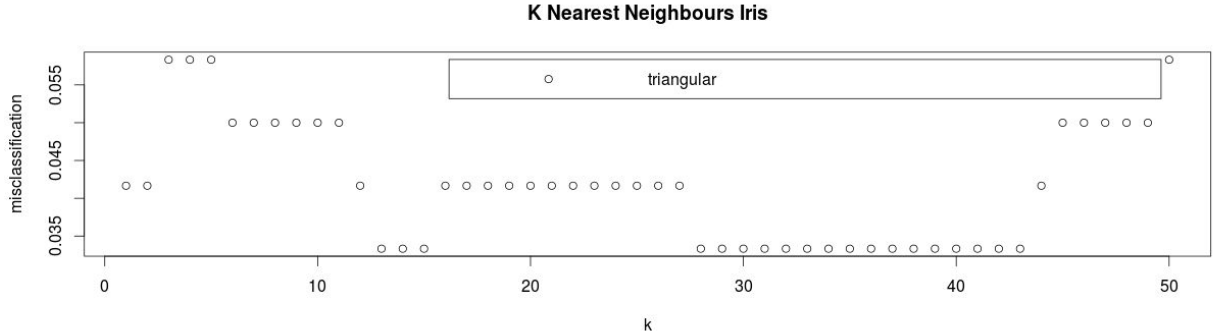
NaiveBase: Confusion Matrix

	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	10	0
virginica	0	0	10
Accuracy:	1		
Recall	1.0000	1.0000	1.0000
Precision	1.0000	1.0000	1.0000
F-Measure	1.0000	1.0000	1.0000



KNN: Confusion Matrix

	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	9	1
virginica	0	1	9
Accuracy :	0.9333		
Recall	1.0000	0.9000	0.9000
Precision	1.0000	0.9000	0.9000
F-Measure	1.0000	0.9	.9

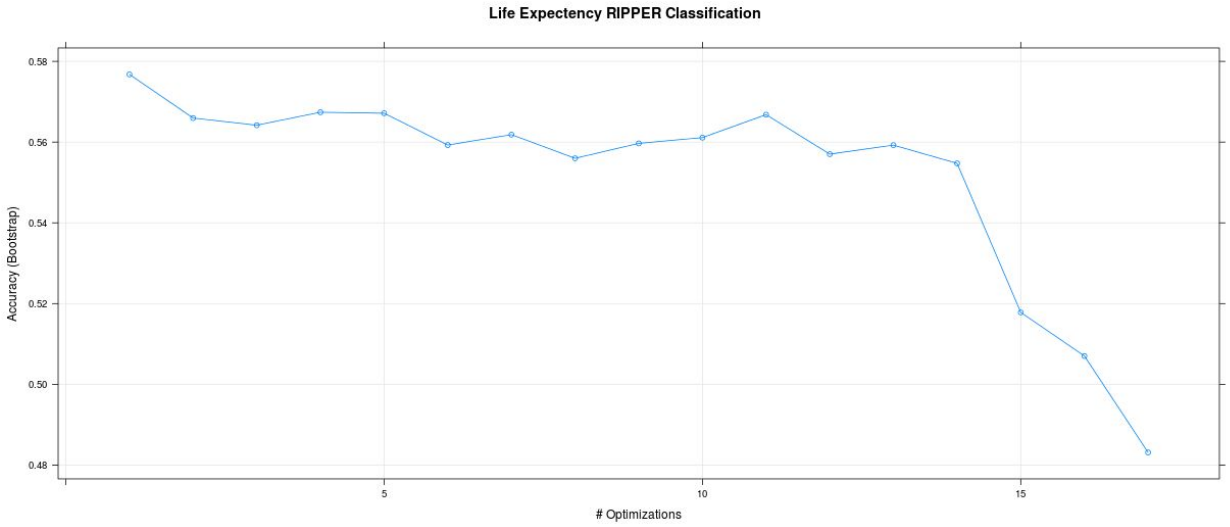


LifeExpectancy Dataset

Desion Tree: (RIPPER)

Confusion Matrix:

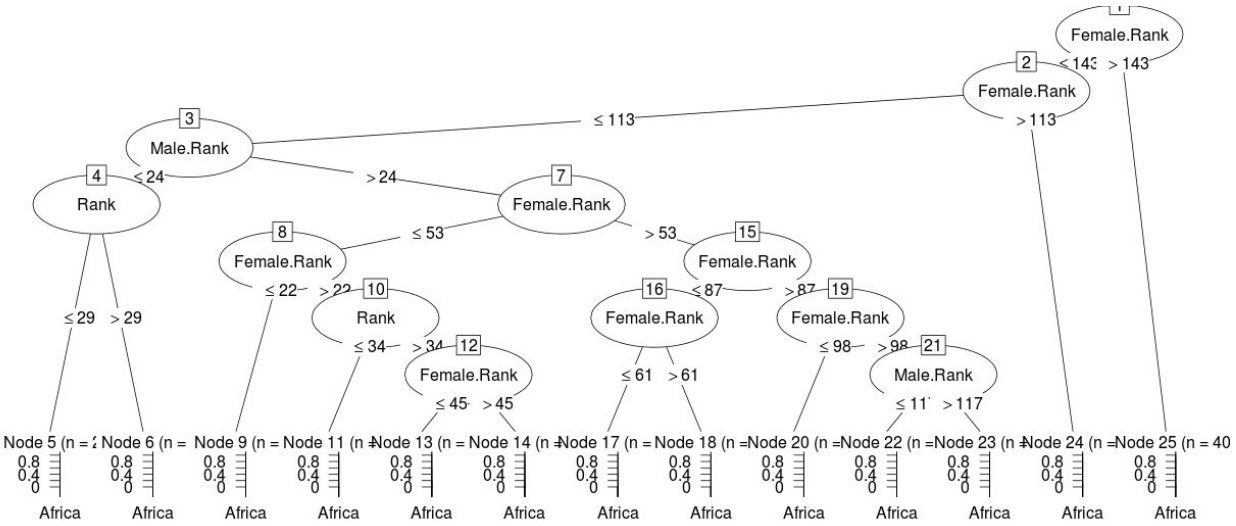
Prediction	Africa	Asia	Europe	North America	South America
Africa	6	1	0	1	0
Asia	4	9	4	3	1
Europe	0	2	4	0	1
North America	0	0	0	0	0
South America	0	0	0	0	0
Accuracy :	0.5278				
Recall	0.6000	0.7500	0.5000	0.0000	0.00000
Precision	0.7500	0.4286	0.5714	NaN	NaN
F-Measure	0.67	0.545	0.533	NaN	NaN



C4.5:

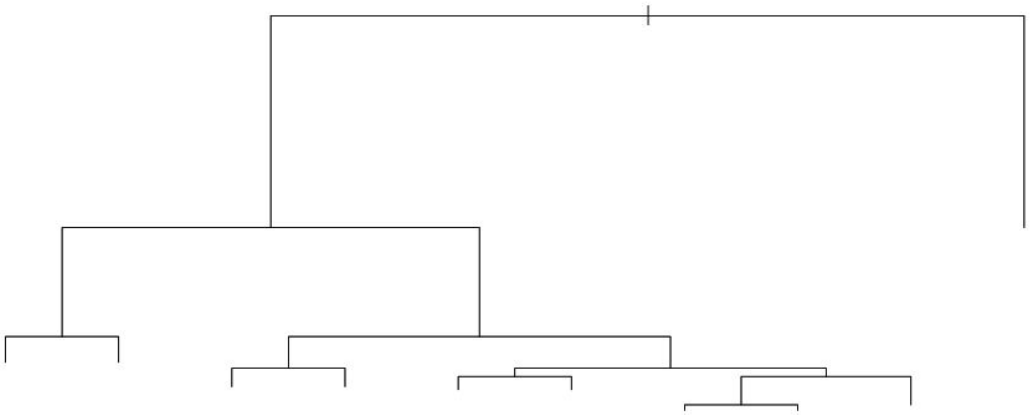
Confusion Matrix

Prediction	Africa	Asia	Europe	North America	South America
Africa	7	1	0	1	0
Asia	3	6	2	2	0
Europe	0	3	3	0	1
North America	0	2	2	1	1
South America	0	0	1	0	0
Accuracy :	0.4722				
Recall	0.7000	0.5000	0.37500	0.25000	0.00000
Precision	0.7778	0.4615	0.42857	0.16667	0.00000
F-Measure	0.74	0.48	0..4	0.20	.000000



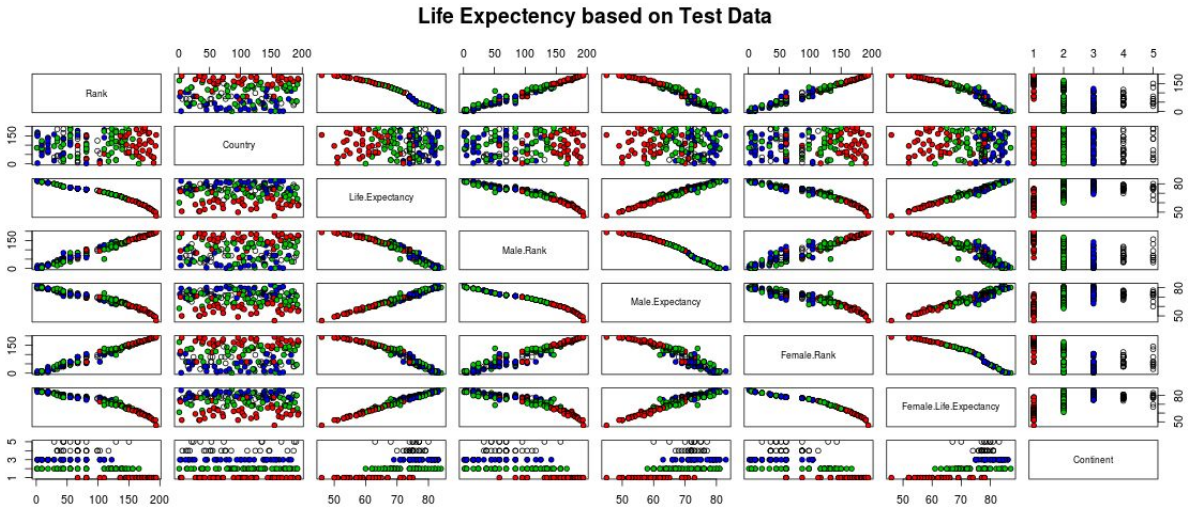
ObliqueTree:

Prediction	Africa	Asia	Europe	North America	South America
Africa	6	0	0	1	0
Asia	4	6	1	2	1
Europe	0	2	6	0	1
North America	0	4	1	1	0
South America	0	0	0	0	0
Accuracy :	0.5278				
Recall	0.6000	0.5000	0.7500	0.25000	0.00000
Precision	0.8571	0.4286	0.6667	0.16667	NaN
F-Measure	0.705	0.461	0.705	0.20	NaN



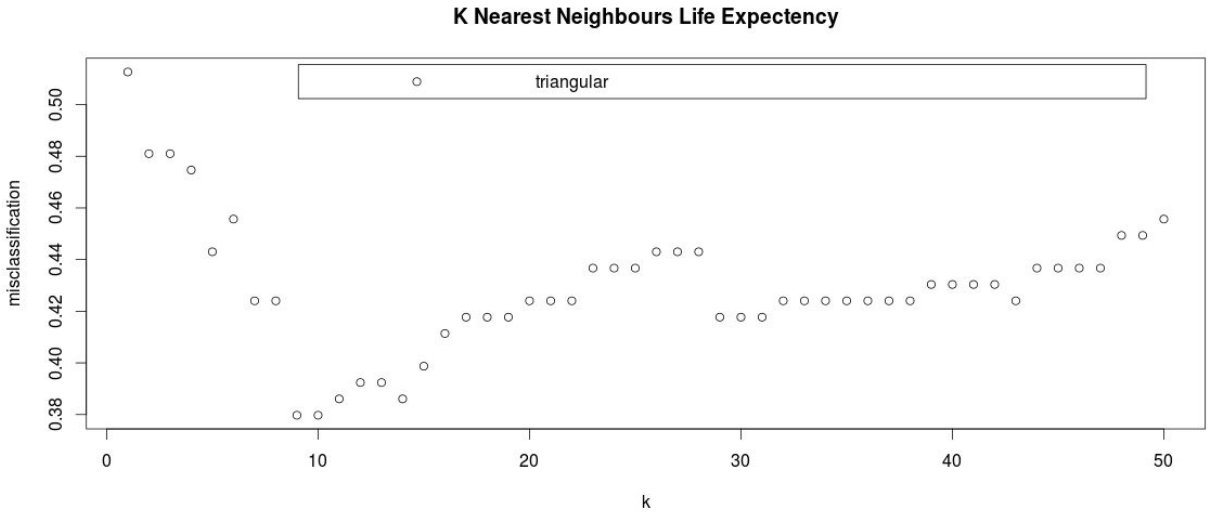
NaiveBase: Confusion Matrix

Prediction	Africa	Asia	Europe	North America	South America
Africa	9	2	0	0	0
Asia	0	4	1	1	0
Europe	0	3	5	1	1
North America	1	3	2	2	1
South America	0	0	0	0	0
Accuracy :	0.5556				
Recall	0.9000	0.3333	0.6250	0.50000	0.00000
Precision	0.8182	0.6667	0.5000	0.22222	NaN
F-Measure	0.857	0.4444	0.5555	0.3076	NaN



K Nearest Neighbours:

Prediction	Africa	Asia	Europe	North America	South America
Africa	8	1	0	2	0
Asia	2	3	0	0	1
Europe	0	4	8	2	1
North America	0	4	0	0	0
South America	0	0	0	0	0
Accuracy :	0.5278				
Recall	0.8000	0.25000	1.0000	0.0000	0.00000
Precision	0.7273	0.50000	0.5333	0.0000	NaN
F-Measure	0.762	0.33	0.695	0.0000	NaN



Conclusion:

In case of the Iris data set Naive Bayes method is having the Highest accuracy.

RIPPER Method	Accuracy : 0.9333
C4.5 Method	Accuracy : 0.9667
Oblique Tree Method	Accuracy : 0.9667
Naive Bayes	Accuracy : 1
KNN Method	Accuracy : 0.9333

In case of the live data (Life Expectancy Dataset) again Naive Bayes is having the highest accuracy.

RIPPER Method	Accuracy : 0.5278
C4.5 Method	Accuracy : 0.4722
Oblique Tree Method	Accuracy : 0.5278
Naive Bayes	Accuracy : 0.5556
KNN Method	Accuracy : 0.5278

The Iris data set is having close to perfect accuracy considering all the methods.