Project-Crime Pattern Analysis

a. The name of the attribute selection method
   Ranker Method with InfoGainAtributeEval attribute evaluator
b. Brief description of the method
   Ranks attribute by their individual evaluation.
c. The names of attributes that were selected by this method

Ranks of attributes: Number of attributes selected	Ranked attributes:
Ranked attributes:	0.294 45 pctKids2Par
0.294 45 pctKids2Par	0.291 44 pct2Par
0.291 44 pct2Par	0.259 47 pct12-17w2Par
0.259 47 pct12-17w2Par	0.258 46 pctKids-4w2Par
0.258 46 pctKids-4w2Par	0.236 51 pctKidsBornNevrMarr
0.236 51 pctKidsBornNevrMarr	0.234 42 pctAllDivorc
Selected attributes: 45,44,47,46,51 : 5	0.233 41 pctFemDivorc
	Selected attributes: 45,44,47,46,5
Ranked attributes:	
0.294 45 pctKids2Par	Ranked attributes:
0.291 44 pct2Par	0.294 45 pctKids2Par
0.259 47 pct12-17w2Par	0.291 44 pct2Par
0.258 46 pctKids-4w2Par	0.259 47 pct12-17w2Par
0.236 51 pctKidsBornNevrMarr	0.258 46 pctKids-4w2Par
0.234 42 pctAllDivorc	0.236 51 pctKidsBornNevrMarr
0.233 41 pctFemDivorc	0.234 42 pctAllDivorc
0.227 28 persPoverty	0.233 41 pctFemDivorc
0.222 29 pctPoverty	0.227 28 persPoverty
0.214 50 kidsBornNevrMarr	0.222 29 pctPoverty
0.21 39 pctMaleDivorc	0.214 50 kidsBornNevrMarr

b. Names of all classifier algorithms you used (including those four specified above).
   1. J48,
   2. RandomForest
   3. SimpleLogistic
   4. MultiLayerPerceptron

c. Test results of all 16 models. You can get the result from Weka's Classifier Output window. Your test result must include all performance measures in the Weka's output window including the confusion matrix.

For 5 attributes

5attributeMLP

=== Summary ===

Correctly Classified Instances	466	77.6667 %
Incorrectly Classified Instances	134	22.3333 %
Kappa statistic	0.4789	
Mean absolute error	0.2852	
Root mean squared error	0.3808	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.601	0.137	0.684	0.601	0.640	0.481	0.853	0.731	0
	0.863	0.399	0.815	0.863	0.838	0.481	0.853	0.907	1
Weighted Avg.	0.777	0.312	0.771	0.777	0.773	0.481	0.853	0.849	

=== Confusion Matrix ===

a	b	<-- classified as
119	79	a = 0
55	347	b = 1

5attributeSimpleLogistic

=== Summary ===

Correctly Classified Instances	471	78.5 %
Incorrectly Classified Instances	129	21.5 %
Kappa statistic	0.4896	
Mean absolute error	0.3012	
Root mean squared error	0.3811	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.581	0.114	0.714	0.581	0.641	0.495	0.853	0.736	0
	0.886	0.419	0.811	0.886	0.847	0.495	0.853	0.910	1
Weighted Avg.	0.785	0.319	0.779	0.785	0.779	0.495	0.853	0.852	

=== Confusion Matrix ===

a	b	<-- classified as
115	83	a = 0
46	356	b = 1

5attributeRandomForest

=== Summary ===

Correctly Classified Instances	469	78.1667 %
Incorrectly Classified Instances	131	21.8333 %
Kappa statistic	0.4979	
Mean absolute error	0.2915	
Root mean squared error	0.3963	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.636	0.147	0.681	0.636	0.658	0.499	0.827	0.690	0
	0.853	0.364	0.827	0.853	0.840	0.499	0.827	0.892	1
Weighted Avg.	0.782	0.292	0.779	0.782	0.780	0.499	0.827	0.826	

=== Confusion Matrix ===

a	b	<-- classified as
126	72	a = 0
59	343	b = 1

5attributeJ48

=== Summary ===

Correctly Classified Instances	456	76 %
Incorrectly Classified Instances	144	24 %
Kappa statistic	0.4708	
Mean absolute error	0.3234	
Root mean squared error	0.4117	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.687	0.204	0.624	0.687	0.654	0.472	0.765	0.552	0
	0.796	0.313	0.838	0.796	0.816	0.472	0.765	0.825	1
Weighted Avg.	0.760	0.277	0.767	0.760	0.763	0.472	0.765	0.735	

=== Confusion Matrix ===

a	b	<-- classified as
136	62	a = 0
82	320	b = 1

For 7 attributes

7attribute J48

=== Summary ===

Correctly Classified Instances	469	78.1667 %
Incorrectly Classified Instances	131	21.8333 %
Kappa statistic	0.4688	
Mean absolute error	0.2878	
Root mean squared error	0.3956	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.530	0.095	0.734	0.530	0.616	0.481	0.803	0.631	0
0.905	0.470	0.796	0.905	0.847	0.481	0.803	0.851	1
Weighted Avg. 0.782	0.346	0.776	0.782	0.771	0.481	0.803	0.779	

=== Confusion Matrix ===

a	b	<-- classified as
105	93	a = 0
38	364	b = 1

7attributeMLP

=== Summary ===

Correctly Classified Instances	464	77.3333 %
Incorrectly Classified Instances	136	22.6667 %
Kappa statistic	0.4821	
Mean absolute error	0.2702	
Root mean squared error	0.391	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.636	0.159	0.663	0.636	0.649	0.482	0.849	0.733	0
0.841	0.364	0.824	0.841	0.833	0.482	0.849	0.910	1
Weighted Avg. 0.773	0.296	0.771	0.773	0.772	0.482	0.849	0.852	

=== Confusion Matrix ===

a	b	<-- classified as
126	72	a = 0
64	338	b = 1

7attributeRandomForest

=== Summary ===

Correctly Classified Instances	429	71.5 %
Incorrectly Classified Instances	171	28.5 %
Kappa statistic	0.3692	
Mean absolute error	0.285	
Root mean squared error	0.5339	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.611	0.234	0.563	0.611	0.586	0.370	0.689	0.472	0
0.766	0.389	0.800	0.766	0.783	0.370	0.689	0.770	1
Weighted Avg. 0.715	0.338	0.722	0.715	0.718	0.370	0.689	0.671	

=== Confusion Matrix ===

a	b	<-- classified as
121	77	a = 0
94	308	b = 1

7attributeSimplelogistic

=== Summary ===

Correctly Classified Instances	473	78.8333 %
Incorrectly Classified Instances	127	21.1667 %
Kappa statistic	0.5132	
Mean absolute error	0.2796	
Root mean squared error	0.3753	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.646	0.142	0.692	0.646	0.668	0.514	0.859	0.743	0
0.858	0.354	0.831	0.858	0.845	0.514	0.859	0.914	1
Weighted Avg. 0.788	0.284	0.785	0.788	0.786	0.514	0.859	0.858	

=== Confusion Matrix ===

a	b	<-- classified as
128	70	a = 0
57	345	b = 1

For 10 attributes

10attributeJ48

=== Summary ===

Correctly Classified Instances	472	78.6667 %
Incorrectly Classified Instances	128	21.3333 %
Kappa statistic	0.5225	
Mean absolute error	0.2788	
Root mean squared error	0.4005	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.697	0.169	0.670	0.697	0.683	0.523	0.789	0.646	0
0.831	0.303	0.848	0.831	0.839	0.523	0.789	0.844	1
Weighted Avg. 0.787	0.259	0.789	0.787	0.788	0.523	0.789	0.778	

=== Confusion Matrix ===

a	b	<-- classified as
138	60	a = 0
68	334	b = 1

10attributeMLP

== Summary ===

Correctly Classified Instances	477	79.5 %
Incorrectly Classified Instances	123	20.5 %
Kappa statistic	0.5286	
Mean absolute error	0.2624	
Root mean squared error	0.376	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.657	0.137	0.703	0.657	0.679	0.529	0.859	0.748	0
0.863	0.343	0.836	0.863	0.849	0.529	0.859	0.906	1
Weighted Avg. 0.795	0.275	0.792	0.795	0.793	0.529	0.859	0.854	

=== Confusion Matrix ===

a	b	<-- classified as
130	68	a = 0
55	347	b = 1

10attributeRandomForest

=== Summary ===

Correctly Classified Instances	437	72.8333 %
Incorrectly Classified Instances	163	27.1667 %
Kappa statistic	0.4077	
Mean absolute error	0.2717	
Root mean squared error	0.5212	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.662	0.239	0.577	0.662	0.616	0.410	0.711	0.493	0
	0.761	0.338	0.820	0.761	0.790	0.410	0.711	0.784	1
Weighted Avg.	0.728	0.306	0.740	0.728	0.733	0.410	0.711	0.688	

=== Confusion Matrix ===

a	b	<-- classified as
131	67	a = 0
96	306	b = 1

10attributeSimpleLogistic

== Summary ===

Correctly Classified Instances	478	79.6667 %
Incorrectly Classified Instances	122	20.3333 %
Kappa statistic	0.5402	
Mean absolute error	0.2686	
Root mean squared error	0.3648	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.692	0.152	0.692	0.692	0.692	0.540	0.871	0.771	0
	0.848	0.308	0.848	0.848	0.848	0.540	0.871	0.917	1
Weighted Avg.	0.797	0.256	0.797	0.797	0.797	0.540	0.871	0.868	

=== Confusion Matrix ===

a	b	<-- classified as
137	61	a = 0
61	341	b = 1

For 15 attributes

15attributeJ48

=== Summary ===

Correctly Classified Instances	465	77.5 %
Incorrectly Classified Instances	135	22.5 %
Kappa statistic	0.5033	
Mean absolute error	0.2775	
Root mean squared error	0.4194	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.707	0.192	0.645	0.707	0.675	0.505	0.773	0.624	0
	0.808	0.293	0.849	0.808	0.828	0.505	0.773	0.832	1
Weighted Avg.	0.775	0.259	0.781	0.775	0.777	0.505	0.773	0.763	

=== Confusion Matrix ===

a	b	<-- classified as
140	58	a = 0
77	325	b = 1

15attributeRandomForest

=== Summary ===

Correctly Classified Instances	485	80.8333 %
Incorrectly Classified Instances	115	19.1667 %
Kappa statistic	0.566	
Mean absolute error	0.2648	
Root mean squared error	0.3717	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.707	0.142	0.711	0.707	0.709	0.566	0.861	0.755	0
	0.858	0.293	0.856	0.858	0.857	0.566	0.861	0.916	1
Weighted Avg.	0.808	0.243	0.808	0.808	0.808	0.566	0.861	0.863	

=== Confusion Matrix ===

a	b	<-- classified as
140	58	a = 0
57	345	b = 1

15attributeSimpleLogistic

=== Summary ===

Correctly Classified Instances	484	80.6667 %
Incorrectly Classified Instances	116	19.3333 %
Kappa statistic	0.5617	
Mean absolute error	0.2662	
Root mean squared error	0.3631	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.702	0.142	0.709	0.702	0.706	0.562	0.874	0.773	0
	0.858	0.298	0.854	0.858	0.856	0.562	0.874	0.920	1
Weighted Avg.	0.807	0.246	0.806	0.807	0.806	0.562	0.874	0.871	

=== Confusion Matrix ===

a	b	<-- classified as
139	59	a = 0
57	345	b = 1

15attributeMLP

=== Summary ===

Correctly Classified Instances	485	80.8333 %
Incorrectly Classified Instances	115	19.1667 %
Kappa statistic	0.5534	
Mean absolute error	0.2242	
Root mean squared error	0.3843	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.652	0.114	0.737	0.652	0.692	0.556	0.859	0.743	0
	0.886	0.348	0.838	0.886	0.861	0.556	0.859	0.914	1
Weighted Avg.	0.808	0.271	0.804	0.808	0.805	0.556	0.859	0.857	

=== Confusion Matrix ===

a	b	<-- classified as
129	69	a = 0
46	356	b = 1

d. The name of the attribute selection method and the classifier algorithm which gave you the *best performance* (e.g., GainRatio attribute selection method with J48 classification algorithm gave me the best performance). You also need to show the list of attributes selected by this attribute selection method.
Answer: As per above results –*Random Forest Classifier using Ranker(InfoGain attribute Eval) Selection* gave the highest accuracy, applied on 15 attributes as follows :-

Ranked attributes:	
0.294	45 pctKids2Par
0.291	44 pct2Par
0.259	47 pct12-17w2Par
0.258	46 pctKids-4w2Par
0.236	51 pctKidsBornNev
0.234	42 pctAllDivorc
0.233	41 pctFemDivorc
0.227	28 persPoverty
0.222	29 pctPoverty
0.214	50 kidsBornNevrM
0.21	39 pctMaleDivorc
0.19	68 pctPersOwnOcc

e. Best test result. This is one of 16 test results you showed in item *c* above. This is the one you obtained by testing your *Best Model* on the *Best Test Dataset*. You can get this result from Weka's Classifier Output window. Your test result must include all performance measures in the Weka's output window including the confusion matrix.

15attributeMLP

=== Summary ===

Correctly Classified Instances	485	80.8333 %
Incorrectly Classified Instances	115	19.1667 %
Kappa statistic	0.5534	
Mean absolute error	0.2242	
Root mean squared error	0.3843	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.652	0.114	0.737	0.652	0.692	0.556	0.859	0.743	0
	0.886	0.348	0.838	0.886	0.861	0.556	0.859	0.914	1
Weighted Avg.	0.808	0.271	0.804	0.808	0.805	0.556	0.859	0.857	

=== Confusion Matrix ===

a	b	<-- classified as
129	69	a = 0
46	356	b = 1

f. Discussion:

(a). Describe what criteria you used when you were choosing your *Best Model*. In other words, you need to justify why you chose that particular model as the best model. If your justification is not based on sound technical criteria, you will lose points.

Answer: Chosen the Random Forest-*for 15 attribute* the best model for its accuracy, time to build and number of correctly classified instances on test data.

Evaluation Criteria		Random Forest
Time to Build Model(in secs)		0.44
Correctly Classified Instances		485
Incorrectly Classified Instances		115
Prediction Accuracy		80.83
Cost	Negative	198
	Positive	402

(b). List five attributes that you think are most relevant to the class attribute. You need to justify why you selected those five attributes.

pctKids2Par

pct2Par

pct12-17w2Par

pctKids-4w2Par

pctKidsBornNevrM

These attributes are most relevant to the class attirbute because it is evident from the above results that whether we select 15,10,7 or 5 attributes for classifying the model these five attributes remain consistent.

(c). What you learned from this project.

This project taught how to split the test-train dataset. Choosing the attributes using attribute selection method and then apply various classification method to determine the Best Model. It taught us to build and test a classifier model using a real-world data using Weka

(d). Any other observations from this project.

It was knowledgeable to observe all the classifier algorithm together in a tab.