APPENDIX

Naïve Bayes: Three methods to tune the parameters were tried using NaiveBayes. The best model was achieved when 'kernel' estimator was used, with 5 incorrect instances.

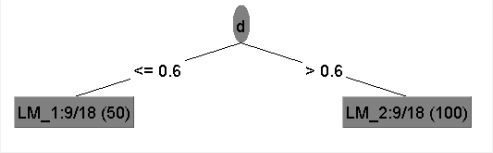
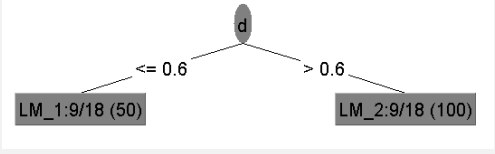
Bayes Classifiers	Percentage Split (66% training)		10 Fold Cross Validation	
NaïveBayes	a b c <-- claessified as 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 48 2 | b = Iris-versicolor 0 4 46 | c = Iris-virginica	Correctly Classified Instances 144 96 % Incorrectly Classified Instances 6 4 %
NaiveBayes (Kernel estimator = true)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 48 2 | b = Iris-versicolor 0 3 47 | c = Iris-virginica	Correctly Classified Instances 145 96.6667 % Incorrectly Classified Instances 5 3.3333 %
NaiveBayes (Supervised Discretization = true)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 44 6 | b = Iris-versicolor 0 5 45 | c = Iris-virginica	Correctly Classified Instances 139 92.6667 % Incorrectly Classified Instances 11 7.3333 %

Decision Tree: Using Weka, A set of decision tree algorithms such as Logistic Model Tree (LMT), Fast decision tree learner (REPTree), Random Forest, Random Tree and J48 (Pruned) were applied with number of iterations on fine tuning to arrive at the best decision tree.

LMT: The default LMT model gave the least incorrect instances (1) at percentage split training & testing.

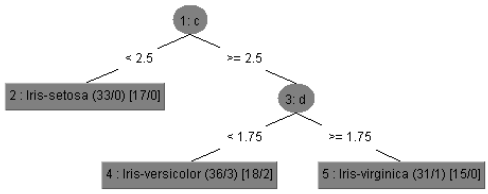
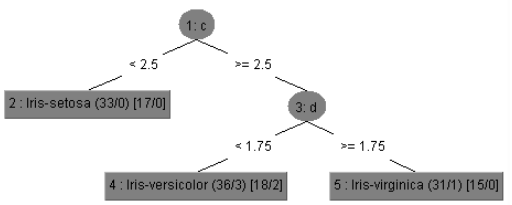
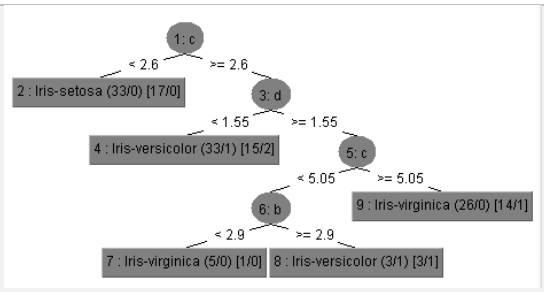
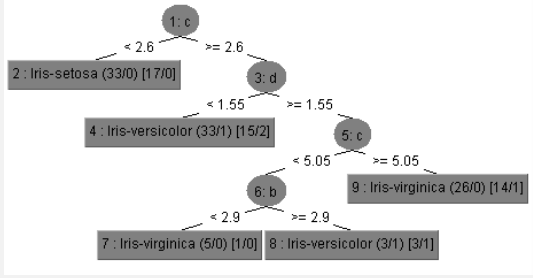
Decision Tree Classifiers	Percentage Split (66% training)		10 Fold Cross Validation	
LMT	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 1 16 | c = Iris-virginica	Correctly Classified Instances 50 98.0392 % Incorrectly Classified Instances 1 1.9608 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 43 7 | b = Iris-versicolor 0 2 48 | c = Iris-virginica	Correctly Classified Instances 141 94 % Incorrectly Classified Instances 9 6 %
LMT (using AIC to stop split)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor	Correctly Classified Instances 49 96.0784 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 47 3 | b = Iris-versicolor	Correctly Classified Instances 145 96.6667 %

	0 2 15 | c = Iris-virginica	Incorrectly Classified Instances 2 3.9216 %	0 2 48 | c = Iris-virginica	Incorrectly Classified Instances 5 3.3333 %

LMT	**Logistic model tree** **: LM_1:18/18 (150)** **Number of Leaves : 1** **Size of the Tree : 1**	**Logistic model tree** **: LM_1:18/18 (150)** **Number of Leaves : 1** **Size of the Tree : 1**
LMT (using AIC to stop split)		

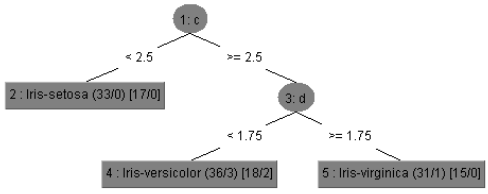
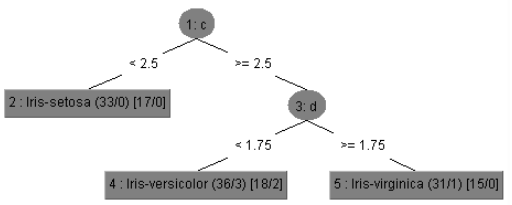
REPTree: With default setting 'seed =1' the model classified with 94% prediction accuracy at 10 Fold Cross Validation. Using 'seed=2' REPTree model classified with 94.7% prediction accuracy at 10 Fold Cross Validation. A difference in seed count caused an extra level to decision tree.

Decision Tree Classifiers	Percentage Split (66% training)		10 Fold Cross Validation	
REPTree (seed =-1)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 15 4 | b = Iris-versicolor 0 0 17 | c = Iris-virginica	Correctly Classified Instances 47 92.1569 % Incorrectly Classified Instances 4 7.8431 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 46 4 | b = Iris-versicolor 0 5 45 | c = Iris-virginica	Correctly Classified Instances 141 94 % Incorrectly Classified Instances 9 6 %
REPTree (seed =2)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 17 2 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 47 92.1569 % Incorrectly Classified Instances 4 7.8431 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 46 4 | b = Iris-versicolor 0 4 46 | c = Iris-virginica	Correctly Classified Instances 142 94.6667 % Incorrectly Classified Instances 8 5.3333 %

REPTree (seed =-1)		
REPTree (seed =2)		

Random Forest:

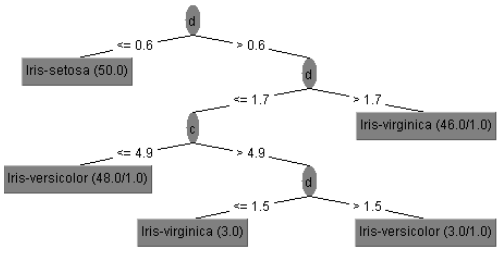
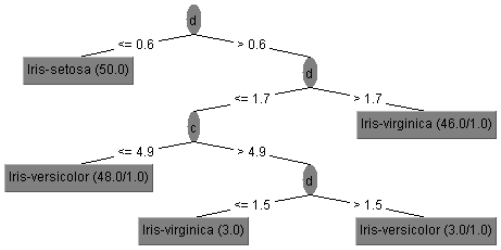
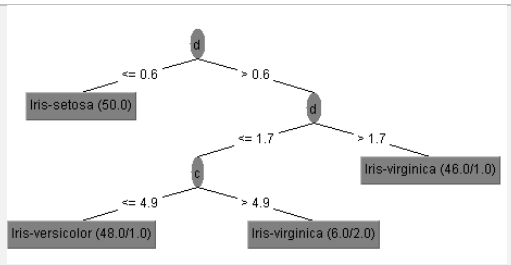
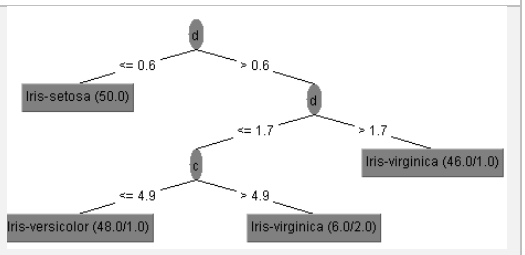
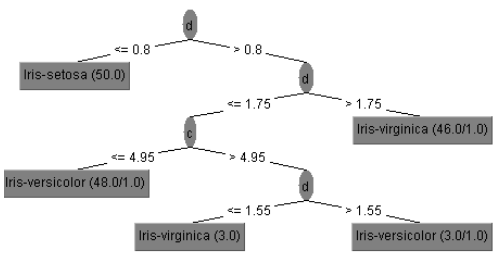
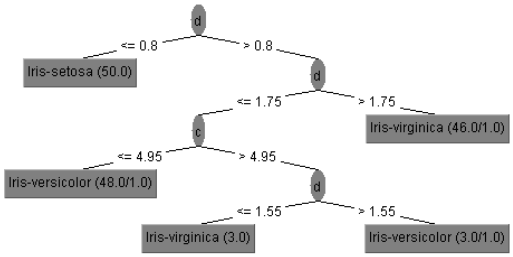
Decision Tree Classifiers	**Percentage Split (66% training)**		**10 Fold Cross Validation**	
Random Forest	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 15 4 | b = Iris-versicolor 0 0 17 | c = Iris-virginica	Correctly Classified Instances 47 92.1569 % Incorrectly Classified Instances 4 7.8431 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 46 4 | b = Iris-versicolor 0 5 45 | c = Iris-virginica	Correctly Classified Instances 141 94 % Incorrectly Classified Instances 9 6 %

Random Forest		

J48: (Pruned C4 Decision tree) the best model was achieved with confidence factor = 0.25 and with 'doNotMakeSplitPointActualValue=true'. There was no impact by 'seed' parameter.

Decision Tree Classifiers	**Percentage Split (66% training)**		**10 Fold Cross Validation**	

J48 (conf = 0.25)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 49 1 0 | a = Iris-setosa 0 47 3 | b = Iris-versicolor 0 2 48 | c = Iris-virginica	Correctly Classified Instances 144 96 % Incorrectly Classified Instances 6 4 %
J48 (conf = 0.1)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginic	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 49 1 0 | a = Iris-setosa 0 46 4 | b = Iris-versicolor 0 2 48 | c = Iris-virginica	Correctly Classified Instances 143 95.3333 % Incorrectly Classified Instances 7 4.6667 %
J48 (conf = 0.25 &-doNotMakeSplitPointActualValue	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 47 3 | b = Iris-versicolor 0 2 48 | c = Iris-virginica	Correctly Classified Instances 145 96.6667 % Incorrectly Classified Instances 5 3.3333 %

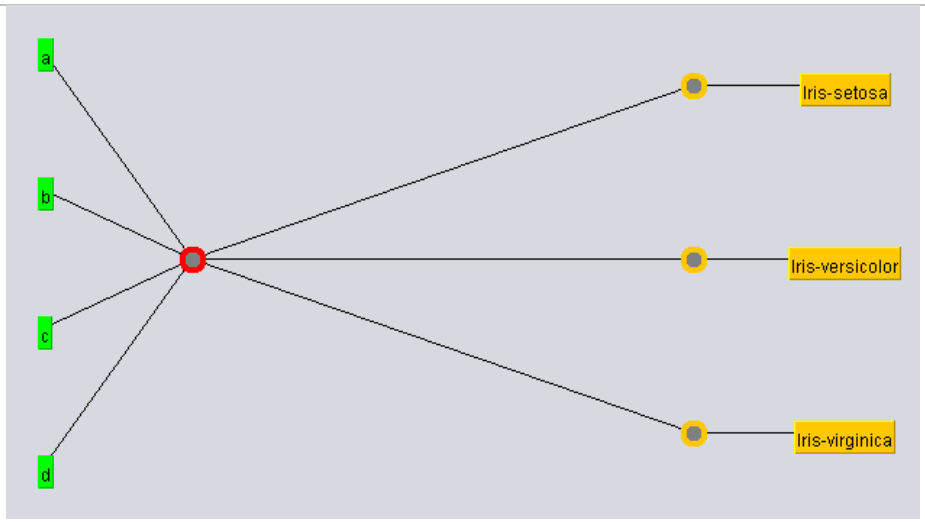
J48 (conf = 0.25)		
J48 (conf = 0.1)		
J48 (conf = 0.25 &-doNotMakeSplitPointActualValue		

k-NearestNeighbour: Using k=12 , which is derived from k=sqrt(N), where N is sample size (150). The 10 Fold cross validation with this configuration gave the results matching the best model.

k-NN Classifiers	Percentage Split (66% training)		10 Fold Cross Validation	
k-NN (k=12) using IBk class	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 49 1 | b = Iris-versicolor 0 4 46 | c = Iris-virginica	Correctly Classified Instances 145 96.6667 % Incorrectly Classified Instances 5 3.3333 %
k-NN (k=12) using IBkLG class	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 48 2 | b = Iris-versicolor 0 4 46 | c = Iris-virginica	Correctly Classified Instances 144 96 % Incorrectly Classified Instances 6 4 %

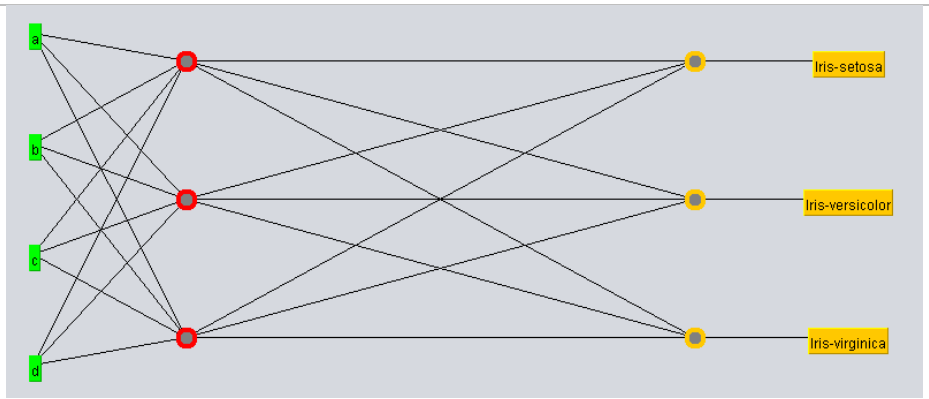
Perceptron (using MultiplePerceptron with h=1): Hidden layer 1 with single neuron, all sigmoid nodes, Learning Rate = 0.3, seed = 0 with normalized attributes match the best model criteria, with 5 incorrect instances at 10 fold cross validation.

Perceptron Classifiers	Percentage Split (66% training)		10 Fold Cross Validation	
Perceptron (seed =0, sigmoid)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 1 16 | c = Iris-virginica	Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 45 5 | b = Iris-versicolor 0 0 50 | c = Iris-virginica	Correctly Classified Instances 145 96.6667 % Incorrectly Classified Instances 5 3.3333 %

Perceptron (seed =0)	

Multiple Layer Neural Network: (**using MultiplePerceptron with h='a' →3).** This model has surpassed the best model ever, with 4 incorrect instances @ 2.667%. (using h='I' → 4), the model is much better than h=a, with only 3 incorrect instances @ 2%.

Multiple Layer Neural Network Classifiers	**Percentage Split (66% training)**		**10 Fold Cross Validation**	
Perceptron (hidden layer = (4+1)/2 → 3)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 1 16 | c = Iris-virginica	Correctly Classified Instances 50 98.0392 % Incorrectly Classified Instances 1 1.9608	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 48 2 | b = Iris-versicolor 0 2 48 | c = Iris-virginica	Correctly Classified Instances 146 97.3333 % Incorrectly Classified Instances 4 2.6667 %
Perceptron (hidden layer = 4)	a b c <-- classified as 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 1 16 | c = Iris-virginica	Correctly Classified Instances 50 98.0392 % Incorrectly Classified Instances 1 1.9608 %	a b c <-- classified as 50 0 0 | a = Iris-setosa 0 48 2 | b = Iris-versicolor 0 1 49 | c = Iris-virginica	Correctly Classified Instances 147 98 % Incorrectly Classified Instances 3 2 %

Perceptron (hidden layer = (4+1)/2 → 3)	
Perceptron (hidden layer = 3)