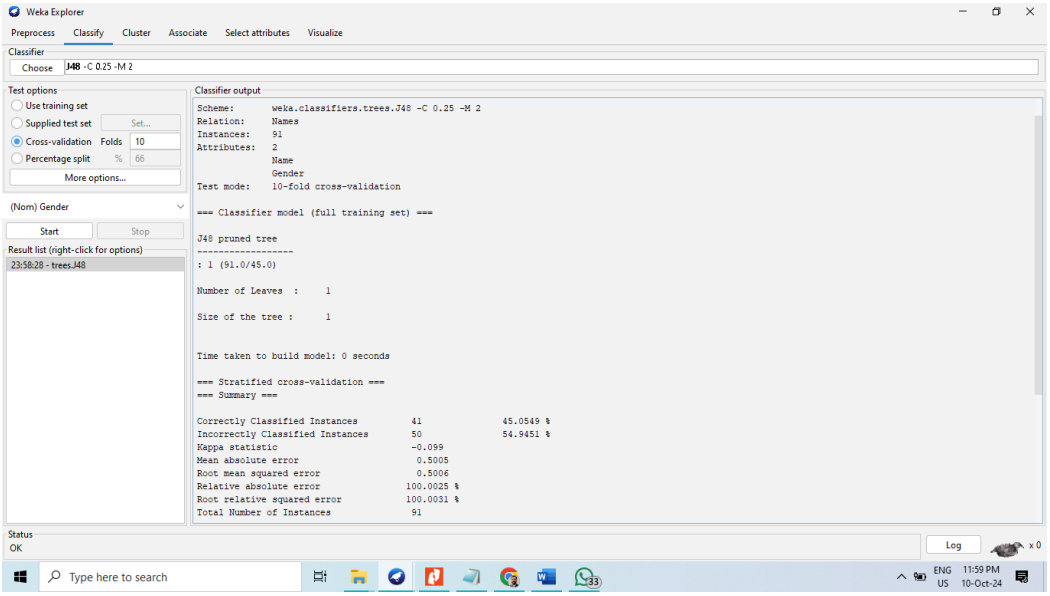
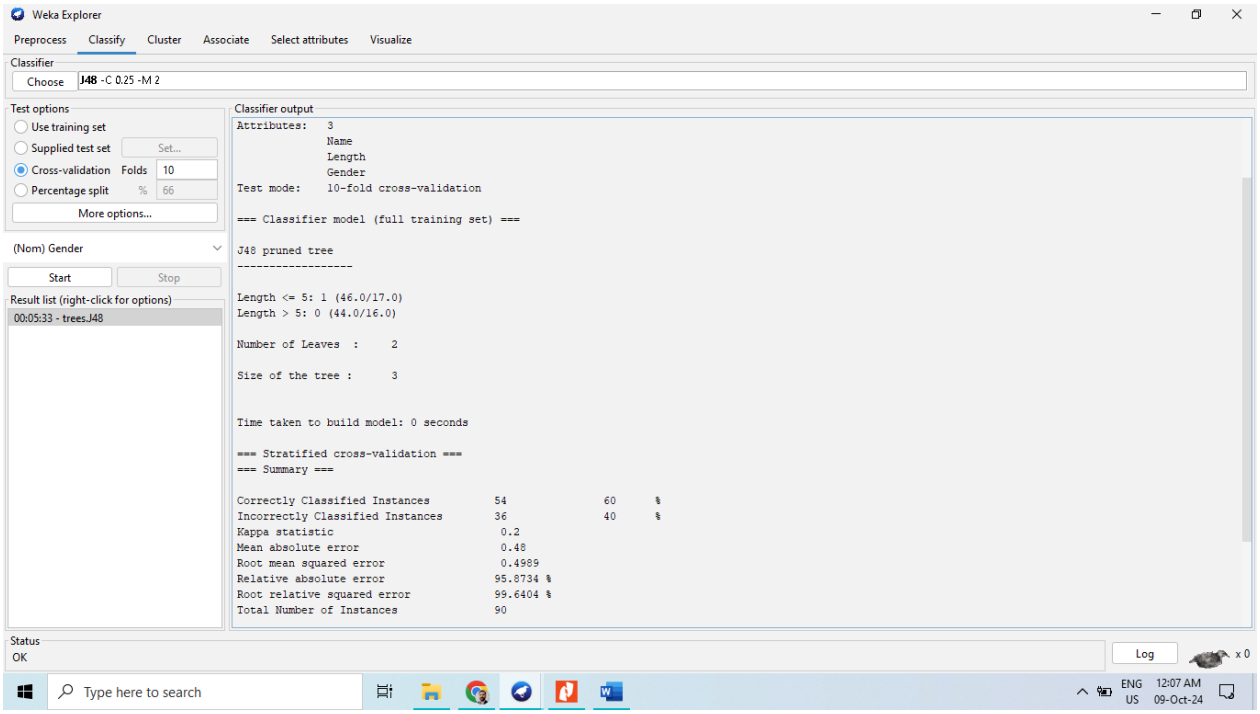
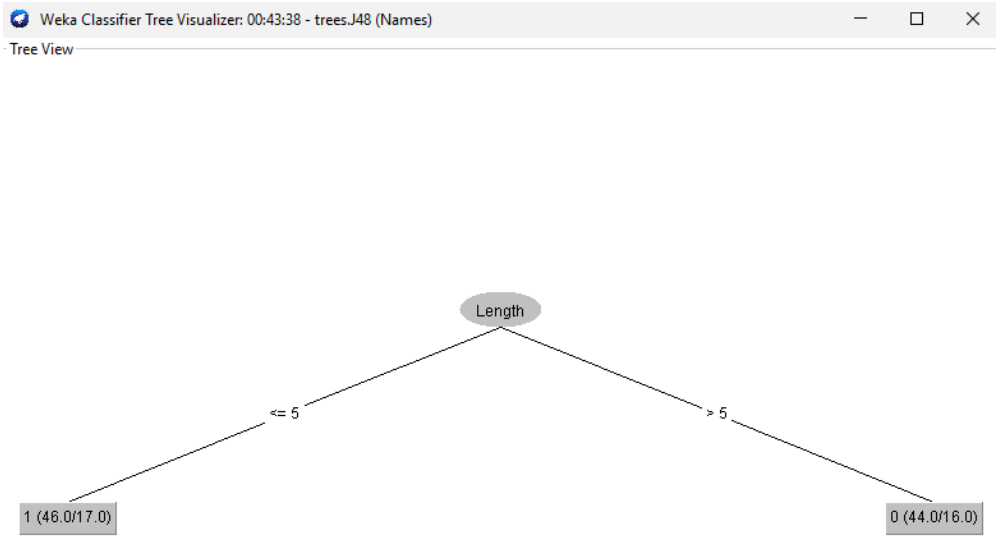
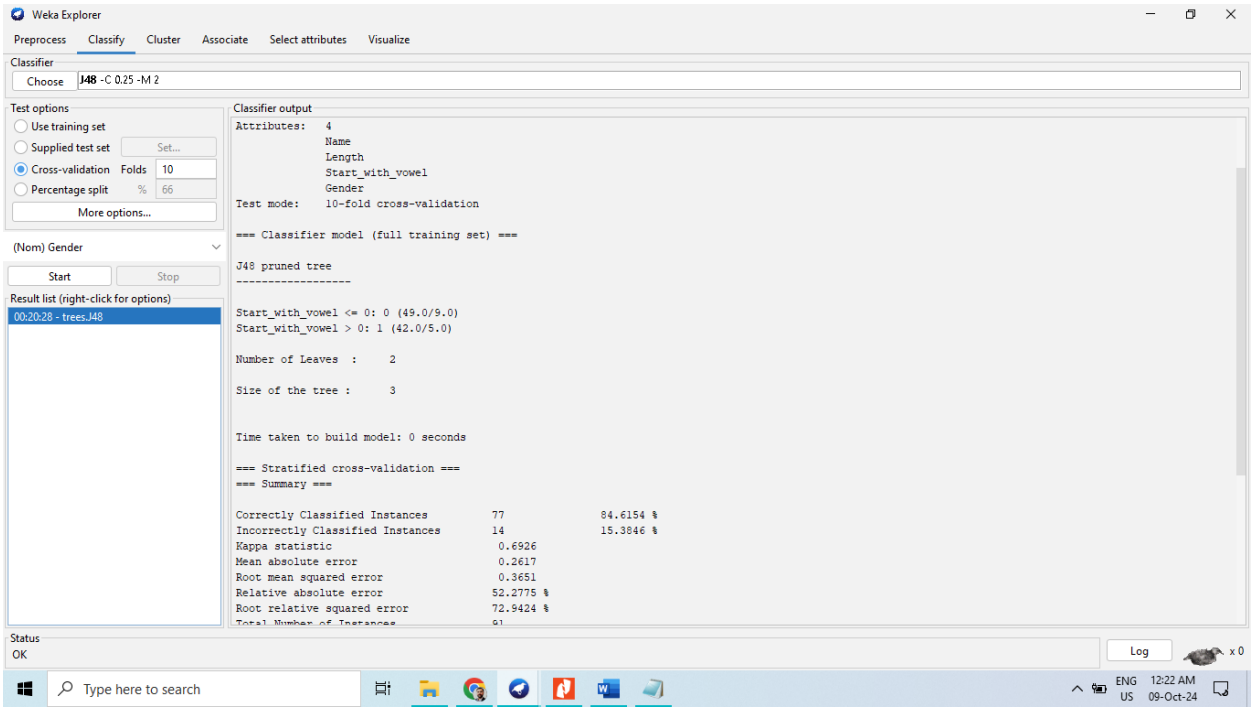
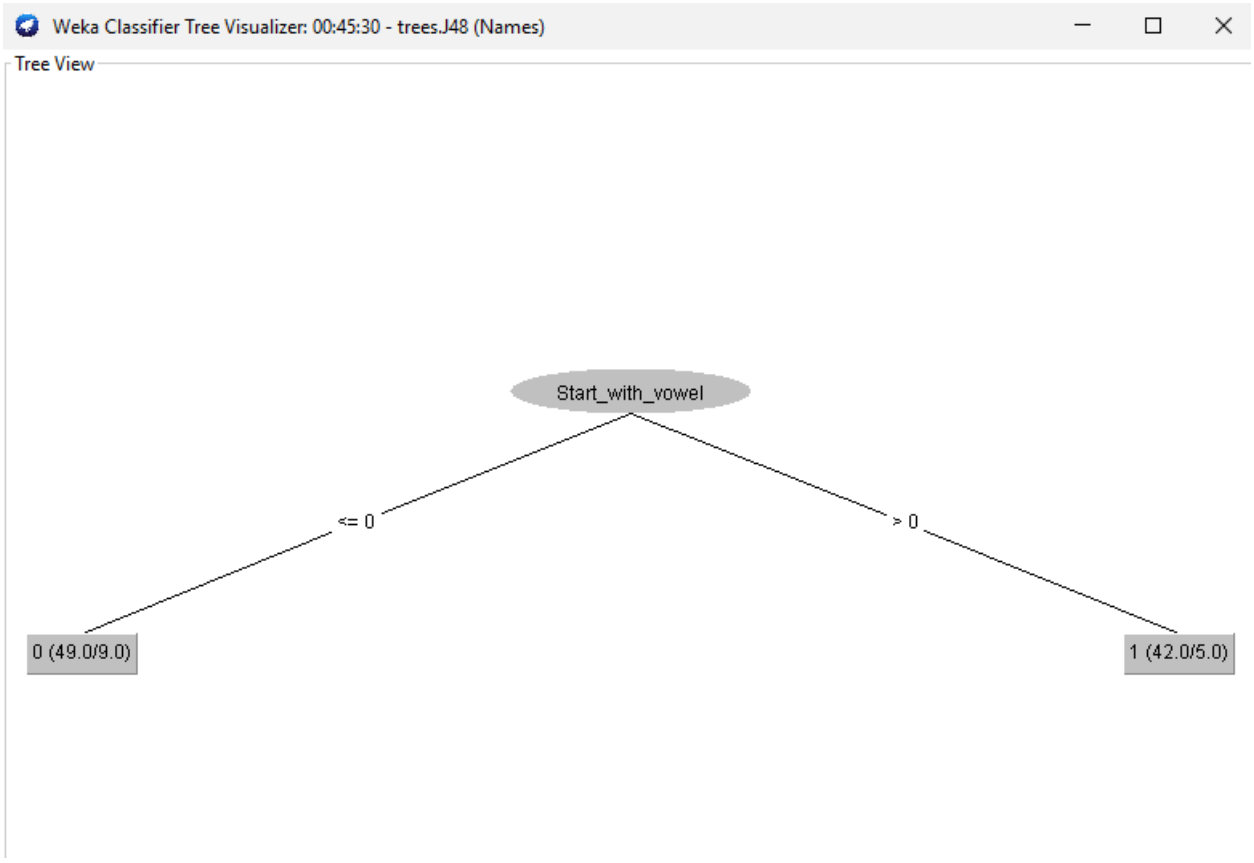
Assignment 02 Machine Learning

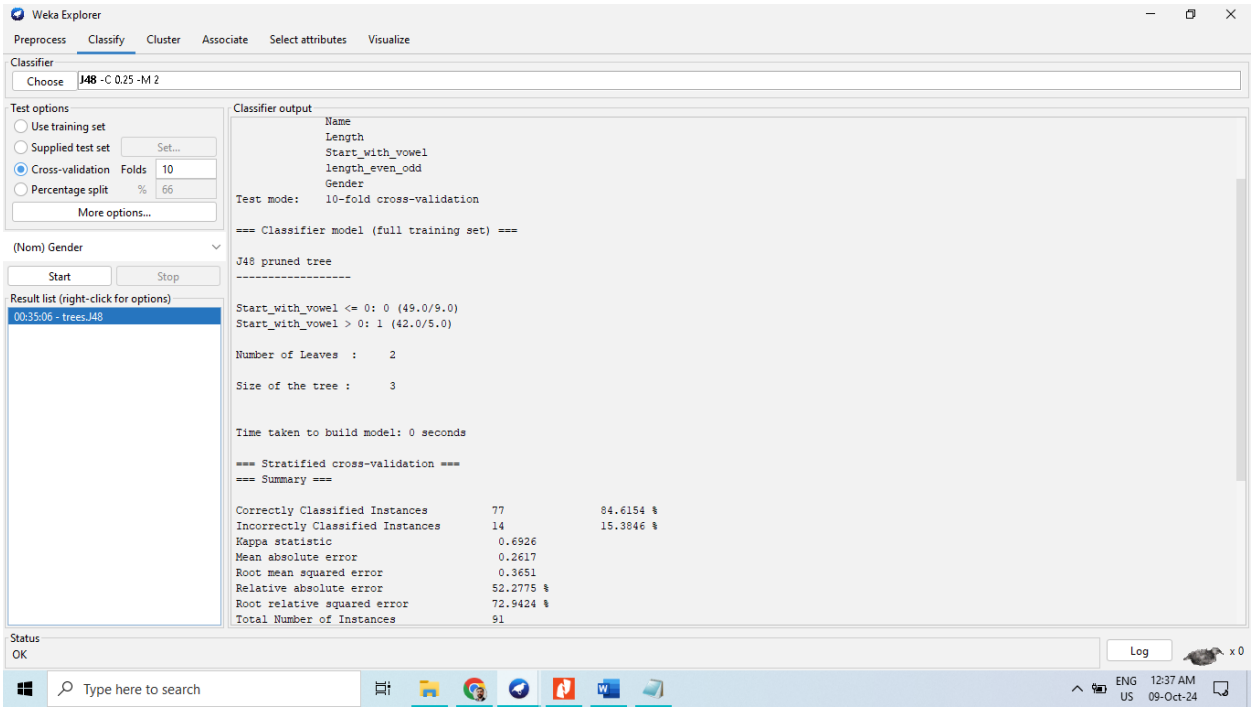


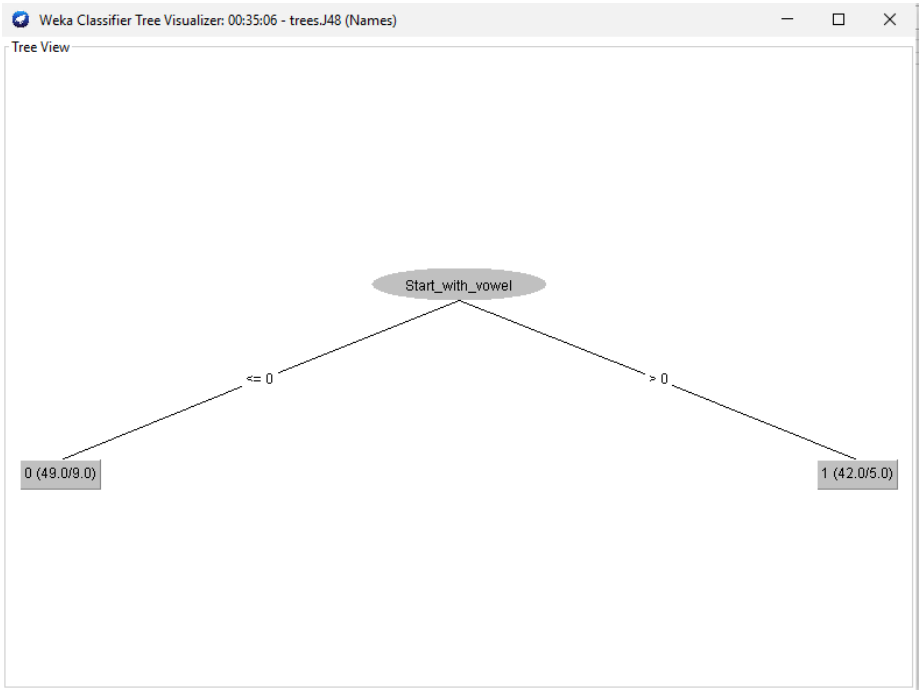

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Gender
Start
Stop
Result list (right-click for options)
00:05:33 - trees.J48
Classifier output
Attributes: 3
Name
Length
Gender
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
Length <= 5: 1 (46.0/17.0)
Length > 5: 0 (44.0/16.0)
Number of Leaves : 2
Size of the tree : 3
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 54 60 %
Incorrectly Classified Instances 36 40 %
Kappa statistic 0.2
Mean absolute error 0.48
Root mean squared error 0.4989
Relative absolute error 95.8734 %
Root relative squared error 99.6404 %
Total Number of Instances 90
Status
OK
Log
x 0
Type here to search
ENG
US
12:07 AM
09-Oct-24


Weka Classifier Tree Visualizer: 00:43:38 - trees.J48 (Names)
Tree View
Length
<= 5
> 5
1 (46.0/17.0)
0 (44.0/16.0)


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Gender
Start
Stop
Result list (right-click for options)
00:20:28 - trees.J48
Classifier output
Attributes: 4
Name
Length
Start_with_vowel
Gender
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
Start_with_vowel <= 0: 0 (49.0/9.0)
Start_with_vowel > 0: 1 (42.0/5.0)
Number of Leaves : 2
Size of the tree : 3
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 77 84.6154 %
Incorrectly Classified Instances 14 15.3846 %
Kappa statistic 0.6926
Mean absolute error 0.2617
Root mean squared error 0.3651
Relative absolute error 52.2775 %
Root relative squared error 72.9424 %
Status
OK
Log
x 0
Type here to search
ENG
US
12:22 AM
09-Oct-24


Weka Classifier Tree Visualizer: 00:45:30 - trees.J48 (Names)
Tree View
Start_with_vowel
<= 0
> 0
0 (49.0/9.0)
1 (42.0/5.0)


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Gender
Start
Stop
Result list (right-click for options)
00:35:06 - trees.J48
Classifier output
Name
Length
Start_with_vowel
length_even_odd
Gender
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
Start_with_vowel <= 0: 0 (49.0/9.0)
Start_with_vowel > 0: 1 (42.0/5.0)
Number of Leaves : 2
Size of the tree : 3
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 77 84.6154 %
Incorrectly Classified Instances 14 15.3846 %
Kappa statistic 0.6926
Mean absolute error 0.2617
Root mean squared error 0.3651
Relative absolute error 52.2775 %
Root relative squared error 72.9424 %
Total Number of Instances 91
Status
OK
Log
x 0
Type here to search
ENG
US
12:37 AM
09-Oct-24


Weka Classifier Tree Visualizer: 00:35:06 - trees.J48 (Names)
Tree View
Start_with_vowel
<= 0
> 0
0 (49.0/9.0)
1 (42.0/5.0)

Running decision tree (J48) algorithm on names data set was quick, and WEKA provided a comprehensive breakdown of the classification results, including the confusion matrix, accuracy, precision, and recall. WEKA explorer also let you preprocess the dataset, you can convert attribute types, add drop attributes, and visualize attributes. Observing the decision tree visually enhanced my understanding of the decision-making process of the model. I also noticed that running J48 strictly requires your dataset's attributes to be of type nominal or numeric. In WEKA, J48 algorithm do not work on string type attribute. Overall, the pipeline offered a clear flow from data preprocessing to model evaluation, making it easier to analyze the strengths and weaknesses of the algorithm. This experience also highlighted the importance of understanding the data and selecting appropriate metrics to assess model performance. The results above clearly indicates that as you increase the number features in a dataset, the accuracy rate of the algorithm also increases. By visualizing the tree, its evident that J48 algorithm chooses most dominant attribute among all the available choices and place it as root node. In case of only single attribute, it selects the one available.