Y.Harika
192325023

CSA1618 DWDM

EXPERIMENT-27

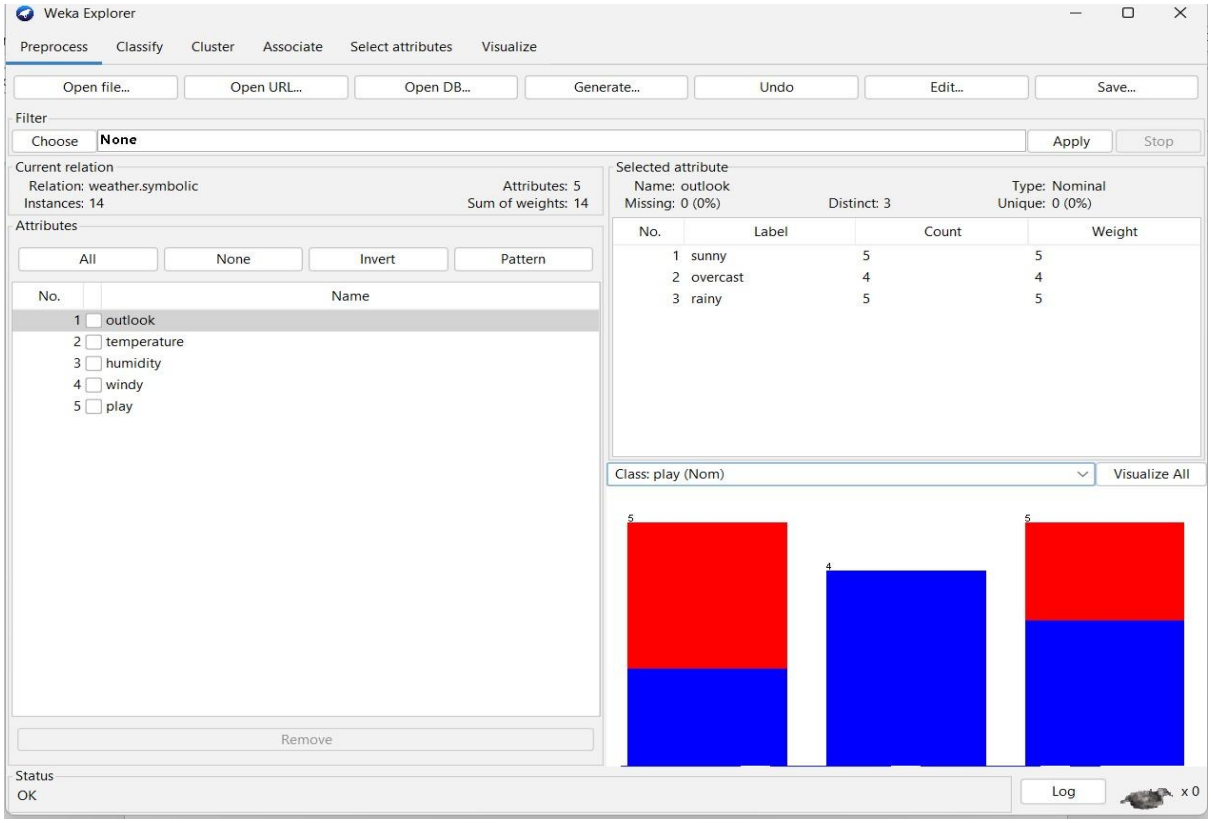
PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORTIHM THROUGH WEKA

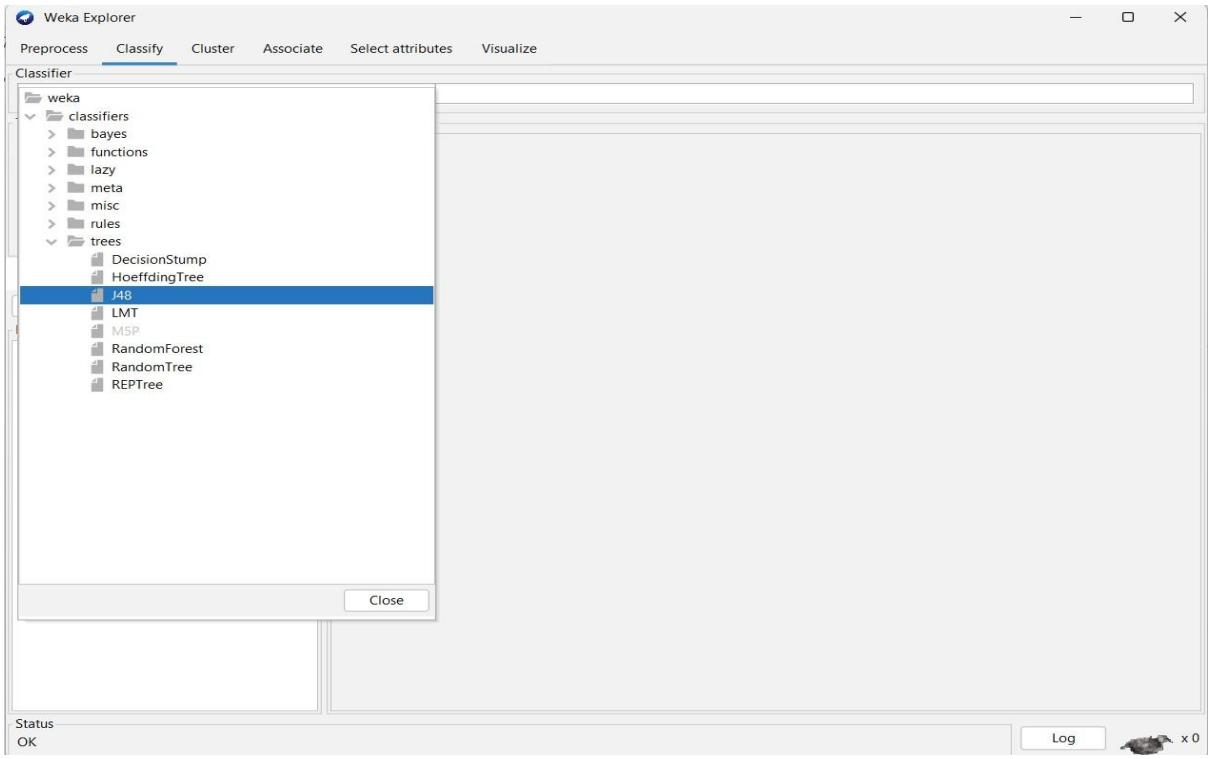
AIM:

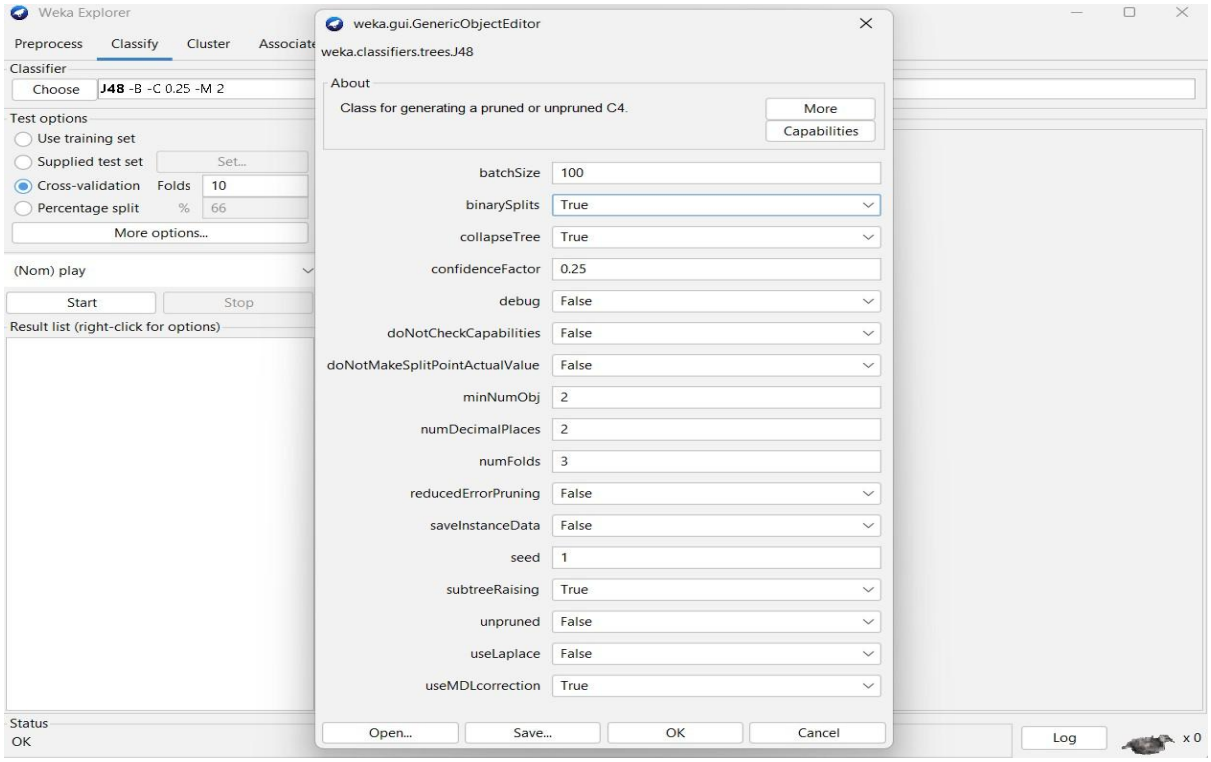
To create prediction of categorical data using decision tree algorithm through weka tool.

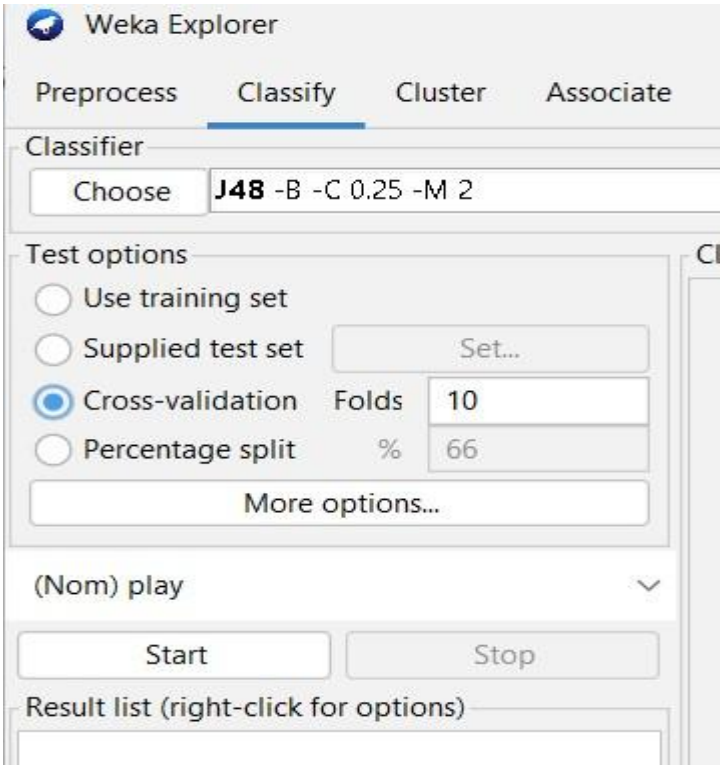
PROCEDURE:

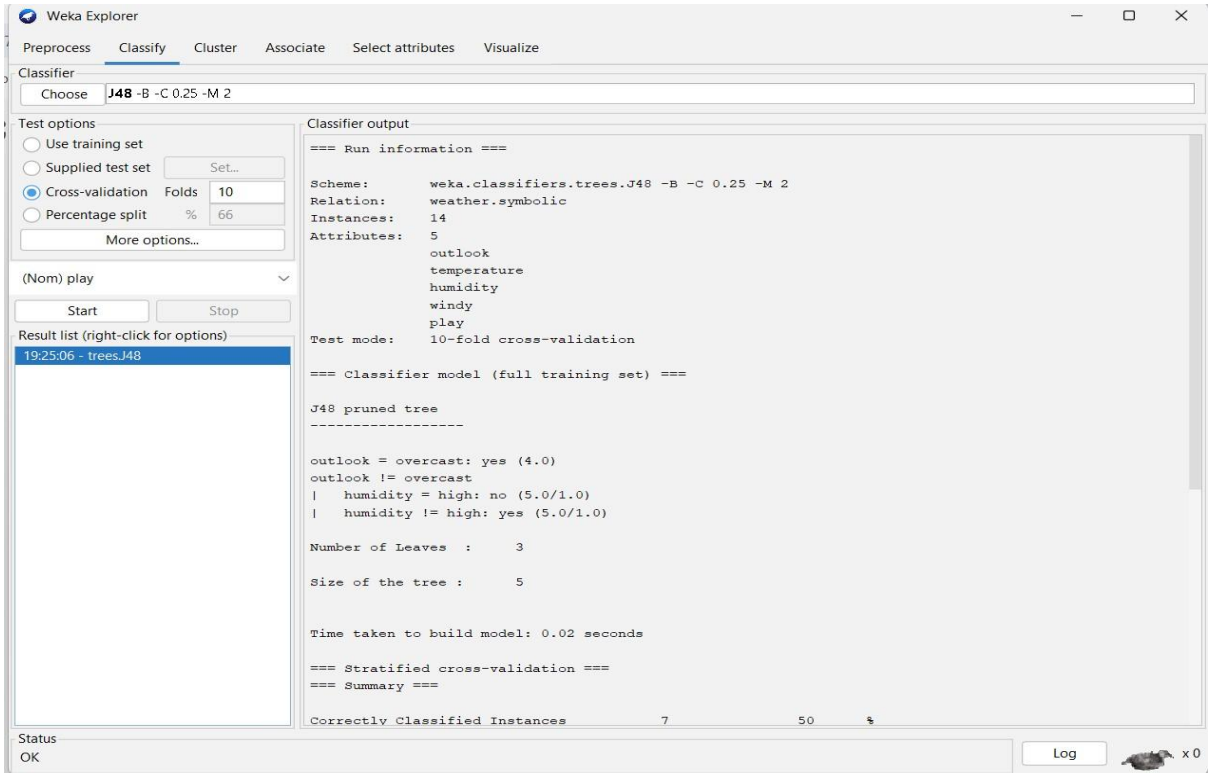
1. Download and install WEKA.
2. Open WEKA and Choose "Explorer" from the main menu.
3. Under Preprocess, Click on the open file button and select the dataset. Ensure that your dataset contains categorical (nominal) attributes.
4. Go to the **Classify** tab, Click **Choose** → Expand the **trees** section → Select **J48** (C4.5 algorithm).
5. Click on J48 to open parameter settings. Modify parameters if needed: Confidence Factor (C): Default is 0.25 (lower values lead to more pruning), MinNumObj (M): Minimum number of instances per leaf (default is 2), Use binary splits (B): Set to True if needed.
6. Under "Test options",the following: Use training set: Trains and tests on the same dataset (not recommended), Cross-validation (default: 10-fold): Splits data into 10 parts for training/testing.
7. Click Start to begin classification.

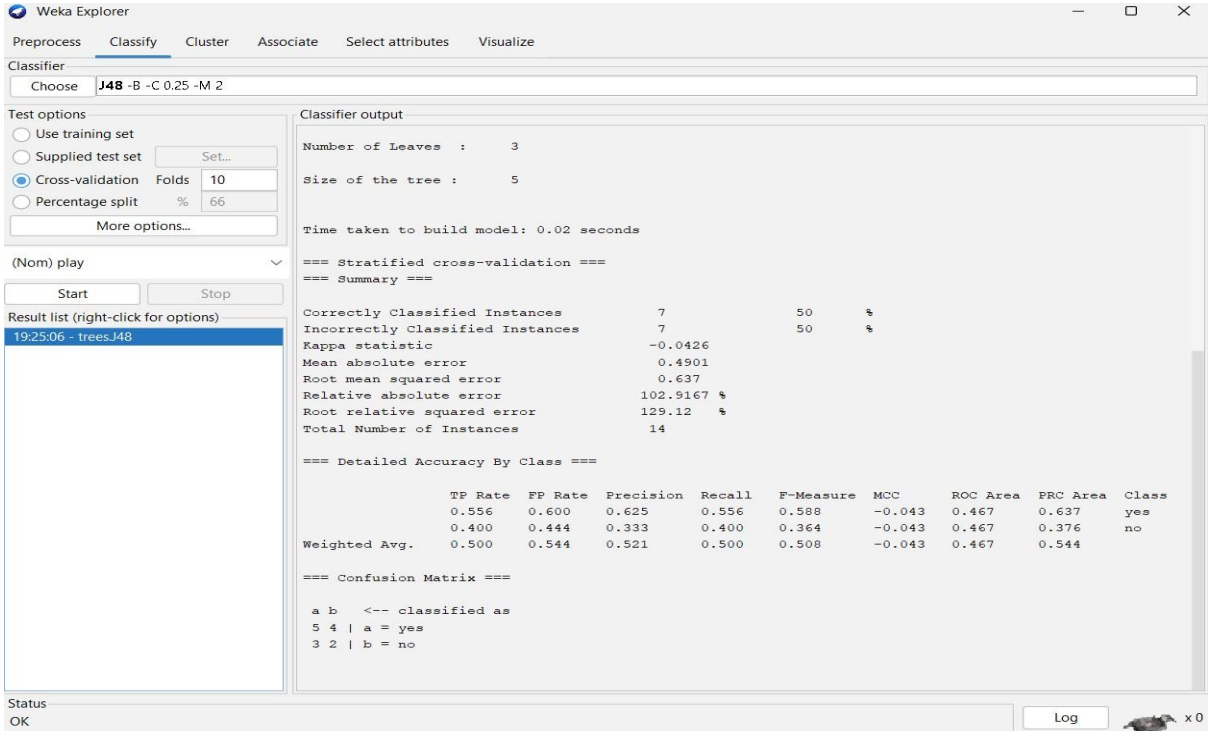

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: weather.symbolic
Instances: 14
Attributes: 5
Sum of weights: 14
Attributes
All
None
Invert
Pattern
No.
Name
1 outlook
2 temperature
3 humidity
4 windy
5 play
Remove
Selected attribute
Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)
No.
Label
Count
Weight
1 sunny 5 5
2 overcast 4 4
3 rainy 5 5
Class: play (Nom)
Visualize All
5
4
5
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
misc
rules
trees
DecisionStump
HoeffdingTree
J48
LMT
M5P
RandomForest
RandomTree
REPTree
Close
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Classifier
Choose
J48 -B -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) play
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.classifiers.trees.J48
About
Class for generating a pruned or unpruned C4.
More
Capabilities
batchSize
100
binarySplits
True
collapseTree
True
confidenceFactor
0.25
debug
False
doNotCheckCapabilities
False
doNotMakeSplitPointActualValue
False
minNumObj
2
numDecimalPlaces
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
seed
1
subtreeRaising
True
unpruned
False
useLaplace
False
useMDLcorrection
True
Open...
Save...
OK
Cancel

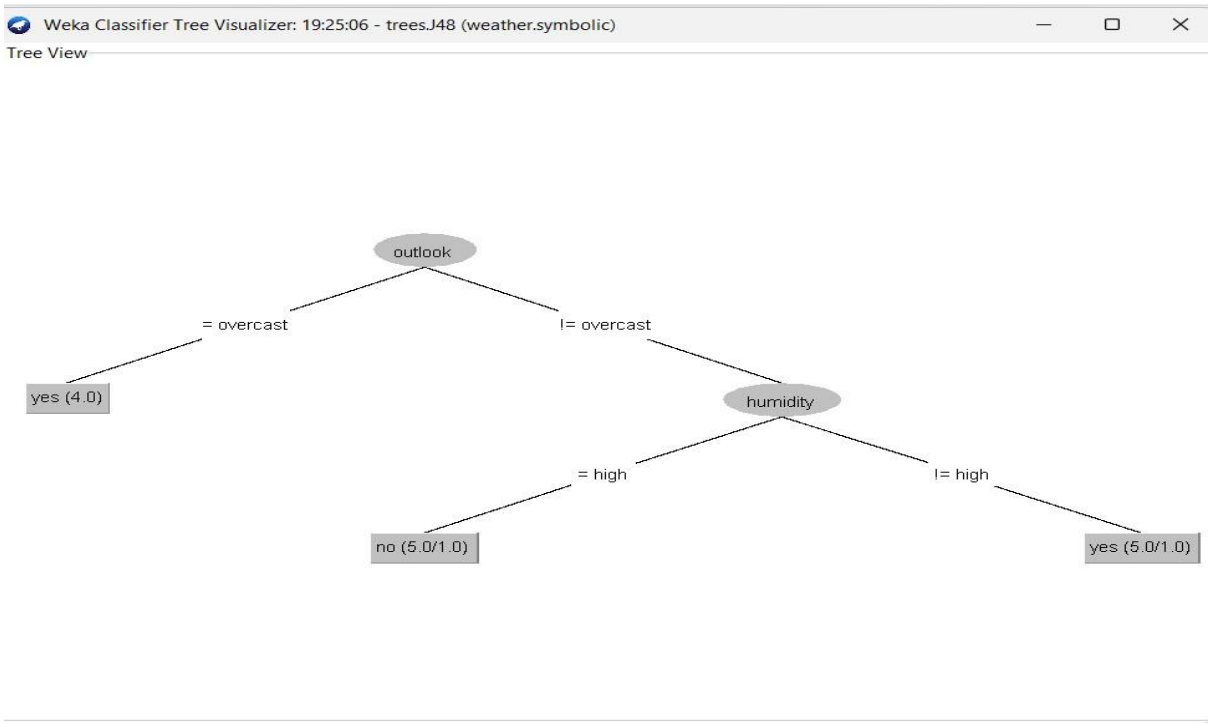

Weka Explorer
Preprocess
Classify
Cluster
Associate
Classifier
Choose
J48 -B -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) play
Start
Stop
Result list (right-click for options)





OBSERVATION:

Cross Validation folds:10



=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no

Root relative squared error 129.12 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	
	0.556	0.600	0.625	0.556	0.588	-0.043	0.467	0.637	yes
	0.400	0.444	0.333	0.400	0.364	-0.043	0.467	0.376	no
Wt Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.467	0.544	

Cross Validation folds:5

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	
0.444	0.600	0.571	0.444	0.500	-0.149	0.400	0.617	yes

	0.400	0.556	0.286	0.400	0.333	-0.149	0.400	0.337	no
Weighted Avg.	0.429	0.584	0.469	0.429	0.440	-0.149	0.400	0.517	

=== Confusion Matrix ===

a b <-- classified as

4 5 | a = yes

3 2 | b = no

RESULT:

Thus, the observations and evaluations done on the dataset are analyzed. The decision tree has been successfully visualized. Various evaluations and comparisons done through the cross validation folds change. Which lead to the change of values in confusion matrix.