S. B. JAIN INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NAGPUR.

Practical No. 4

Aim: Construct the Decision Trees for Weather, Customer and Location data set by changing the value of Binary Splits for Classification and Prediction.

Name of Student: Shrutika Pradeep Bagdi

Roll No.:   CS22130

Semester/Year: 6th / 3rd

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Construct the Decision Trees for Weather, Customer and Location data set by changing the value of Binary Splits for Classification and Prediction

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- To map of the possible outcomes of a series of related choices.
- It allows an individual or organization to weigh possible actions against one another based on their costs, probabilities, and benefits.
- To split the data in such a way that at the end we have different groups of data which has more similarity and less randomness/impurity

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement:

Software Requirement:

Weka Tool-3.6.9

THEORY:

A DECISION TREE:-

A decision tree is a map of the possible outcomes of a series of related choices. It allows an individual or organization to weigh possible actions against one another based on their costs, probabilities, and benefits. They can be used either to drive informal discussion or to map out an algorithm that predicts the best choice mathematically**.**

A decision tree typically starts with a single node, which branches into possible outcomes. Each of those outcomes leads to additional nodes, which branch off into other possibilities. This gives it a treelike shape.

There are three different types of nodes: chance nodes, decision nodes, and end nodes. A chance node, represented by a circle, shows the probabilities of certain results. A decision node, represented by a square, shows a decision to be made, and an end node shows the final outcome of a decision path.

DECISION TREES IN DATA MINING

A decision tree can also be used to help build automated predictive models, which have applications in machine learning, data mining, and statistics. Known as decision tree learning, this method takes into account observations about an item to predict that item's value.

In these decision trees, nodes represent data rather than decisions. This type of tree is also known as a classification tree. Each branch contains a set of attributes, or classification rules, that are associated with a particular class label, which is found at the end of the branch.

These rules, also known as decision rules, can be expressed in an if-then clause, with each decision or data value forming a clause, such that, for instance, "if conditions 1, 2 and 3 are fulfilled, then outcome x will be the result with y certainty."

For increased accuracy, sometimes multiple trees are used together in ensemble methods:

- **Bagging** creates multiple trees by resampling the source data, then has those trees vote to reach consensus.

- **A Random Forest** classifier consists of multiple trees designed to increase the classification rate

- **Boosted trees** that can be used for regression and classification trees.

- The trees in a **Rotation Forest** are all trained by using PCA (principal component analysis) on a random portion of the data

Description:

Classification & Prediction: Classification is the process for finding a model that describes the data values and concepts for the purpose of Prediction.

Decision Tree: A decision Tree is a classification scheme to generate a tree consisting of root node, internal nodes and external nodes. Root nodes representing the attributes. Internal nodes are also the attributes. External nodes are the classes and each branch represents the values of the attributes Decision Tree also contains set of rules for a given data set; there are two subsets in Decision Tree. One is a Training data set and second one is a Testing data set. Training data set is previously classified data. Testing data set is newly generated data.

Procedure:

1) Open Start -> Programs -> Accessories -> Notepad

2.a) Type the following training data set with the help of Notepad for **Weather data set.**

@relation weather

@attribute outlook {sunny,rainy,overcast}

@attribute temparature numeric

@attribute humidity numeric

@attribute windy {true,false}

@attribute play {yes,no}

@data

sunny,85.0,85.0,false,no

overcast,80.0,90.0,true,no

sunny,83.0,86.0,false,yes

rainy,70.0,86.0,false,yes

rainy,68.0,80.0,false,yes

rainy,65.0,70.0,true,no

overcast,64.0,65.0,false,yes

sunny,72.0,95.0,true,no

sunny,69.0,70.0,false,yes

rainy,75.0,80.0,false,yes

overcast,78.0,75.0,true,yes

sunny,82.0,85.0,false,no

sunny,79.0,60.0,true,no

rainy,65.0,70.0,false,no

2.b) Type the following training data set with the help of Notepad for **Customer data set.**

@relation customer

@attribute name {x,y,z,u,v,l,w,q,r,n}

@attribute age {youth,middle,senior}

@attribute income {high,medium,low}

@attribute class {A,B}

@data

x,youth,high,A

z,middle,high,A

v,senior,high,A

w,youth,high,A

r,middle,high,A

n,senior,high,A

y,youth,low,A

u,middle,low,A

l,senior,low,A

q,youth,low,A

2.c) Type the following training data set with the help of Notepad for **Location data set.**

@relation location

@attribute age {21,24,25}

@attribute location {hyd,blr,kdp}

@data

21,hyd

21,hyd

24,blr

24,blr

24,blr

24,blr

21,hyd

25,kdp

25,kdp

25,kdp

Procedure for Decision Trees:

1) Open Start , Programs , Weka-3-6.9
2) Open explorer.
3) Click on open file and select wether.arff / customer.arff / location.arff
4) Select Classifier option on the top of the Menu bar.

5) Select Choose button and click on Tree option.
6) Click on J48.
7) Click on Start button and output will be displayed on the right side of the window.
8) Select the result list and right click on result list and select Visualize Tree option.
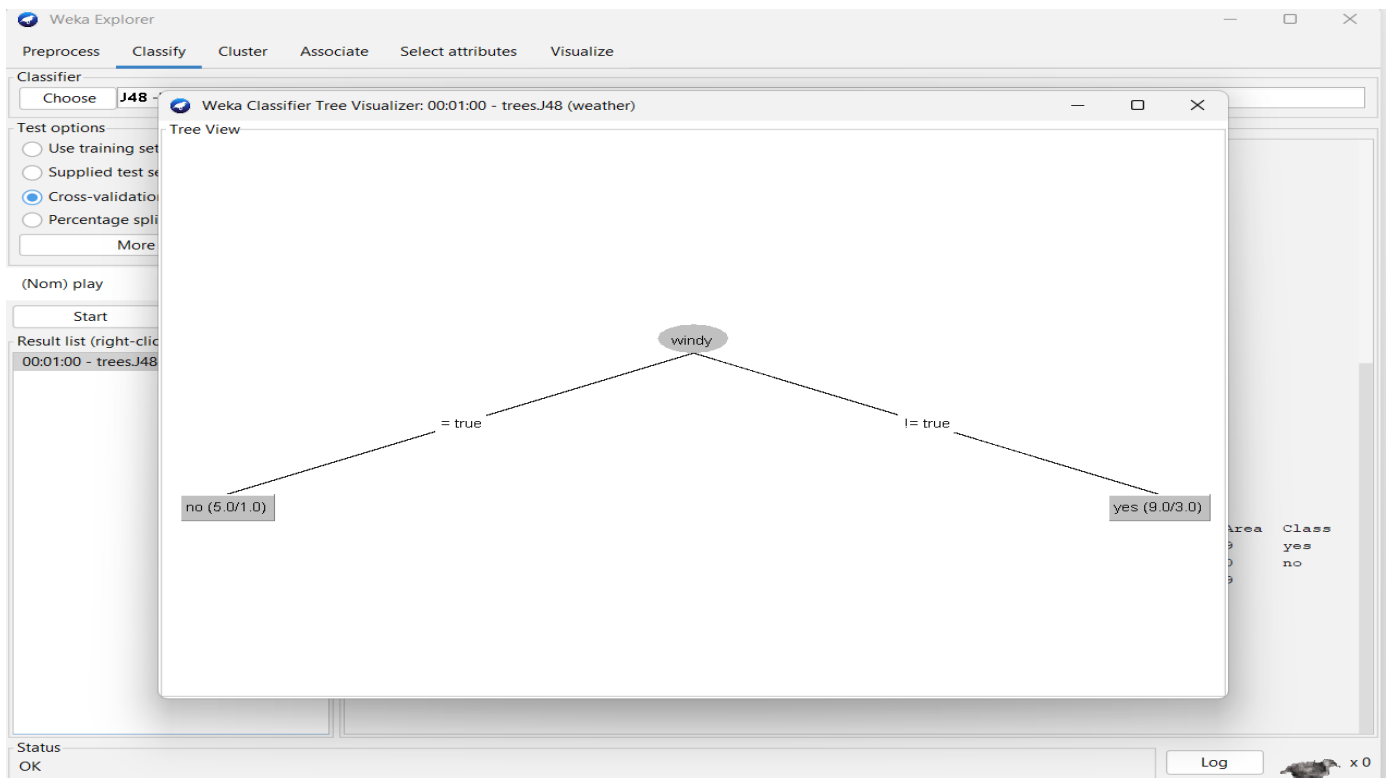9) Then Decision Tree will be displayed on new window

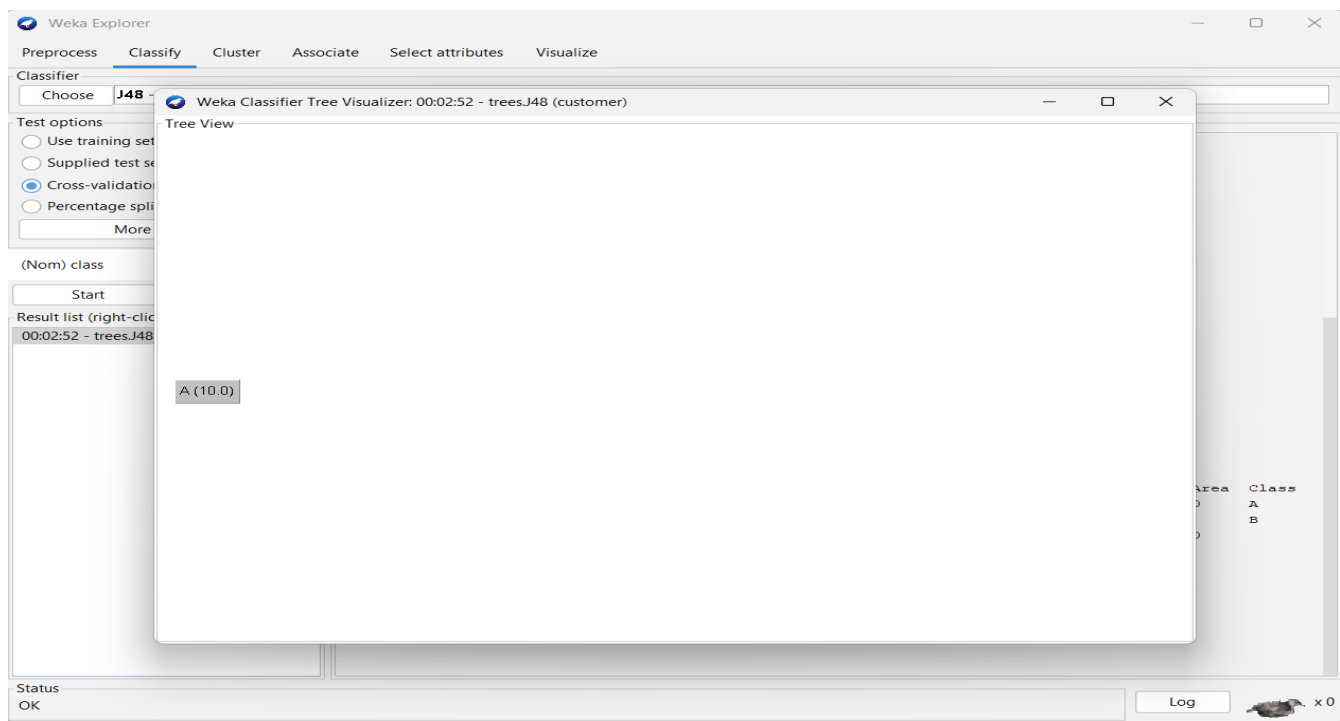
OUTPUT (SCREENSHOTS):
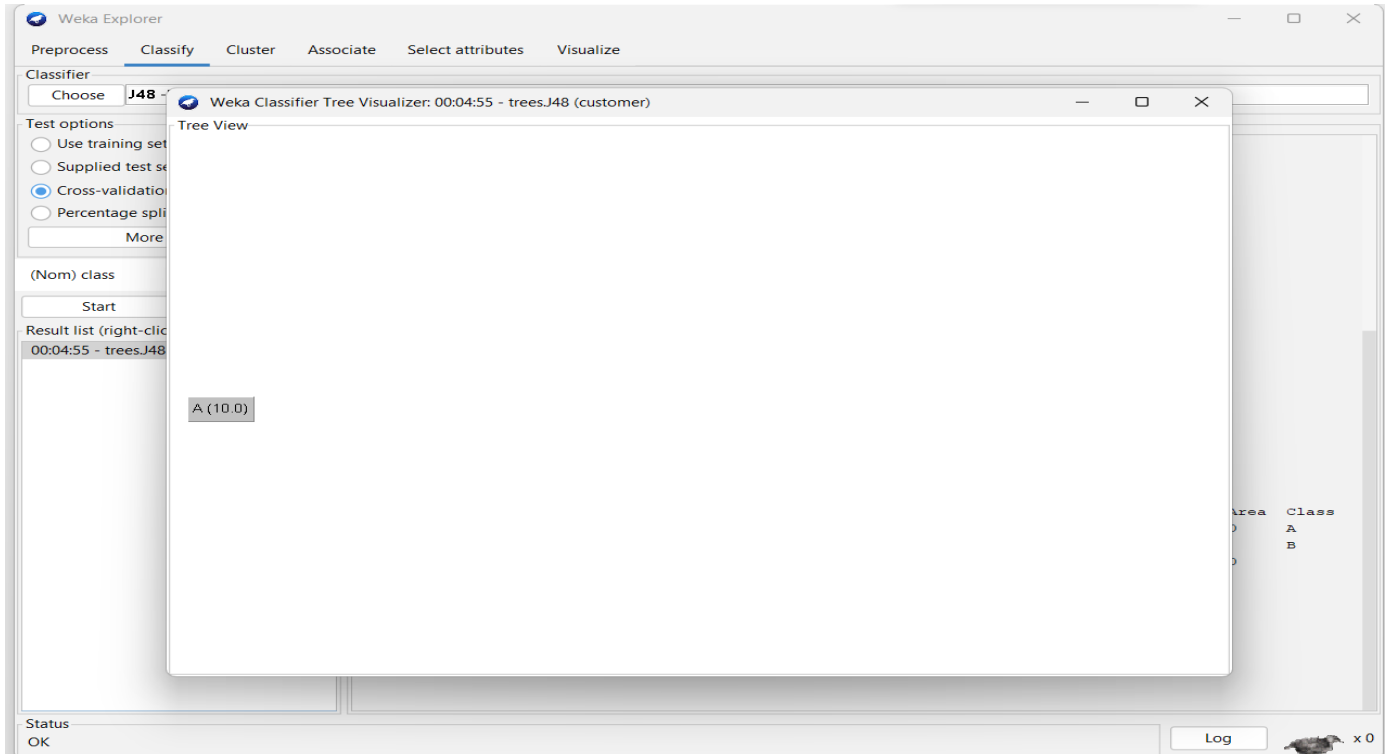
Weather:
BinarySplit = False:
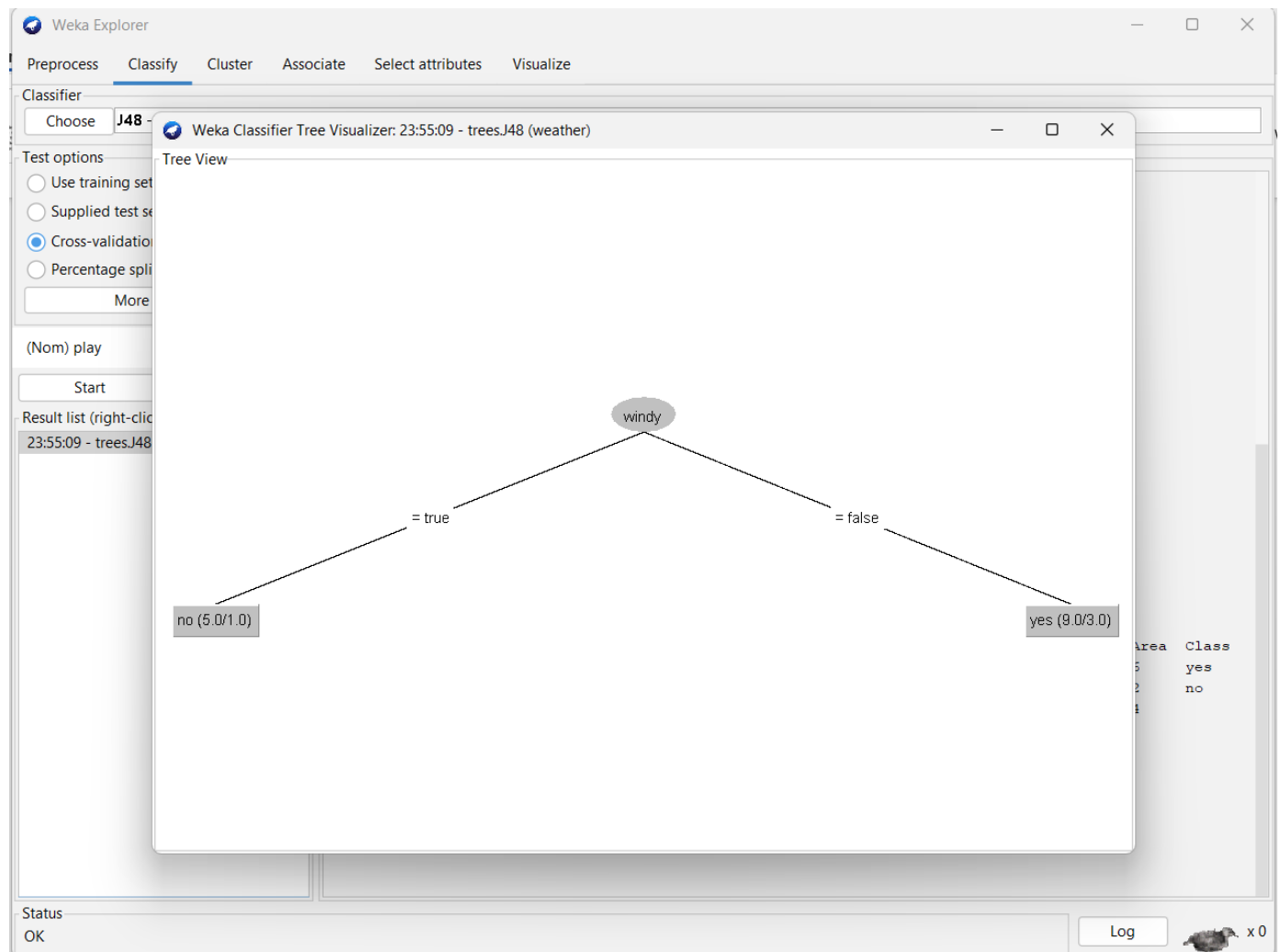
BinarySplit = True:

Customer:

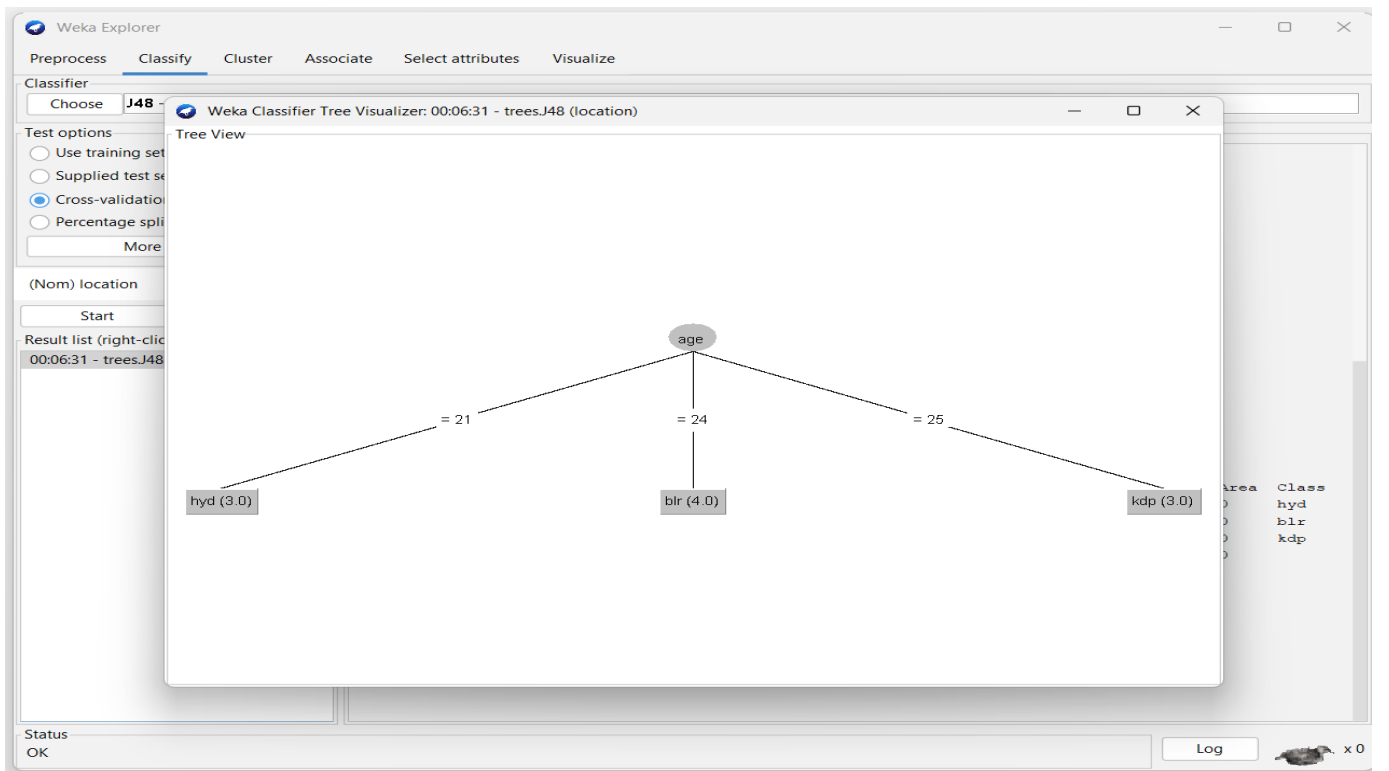**BinarySplit = False**
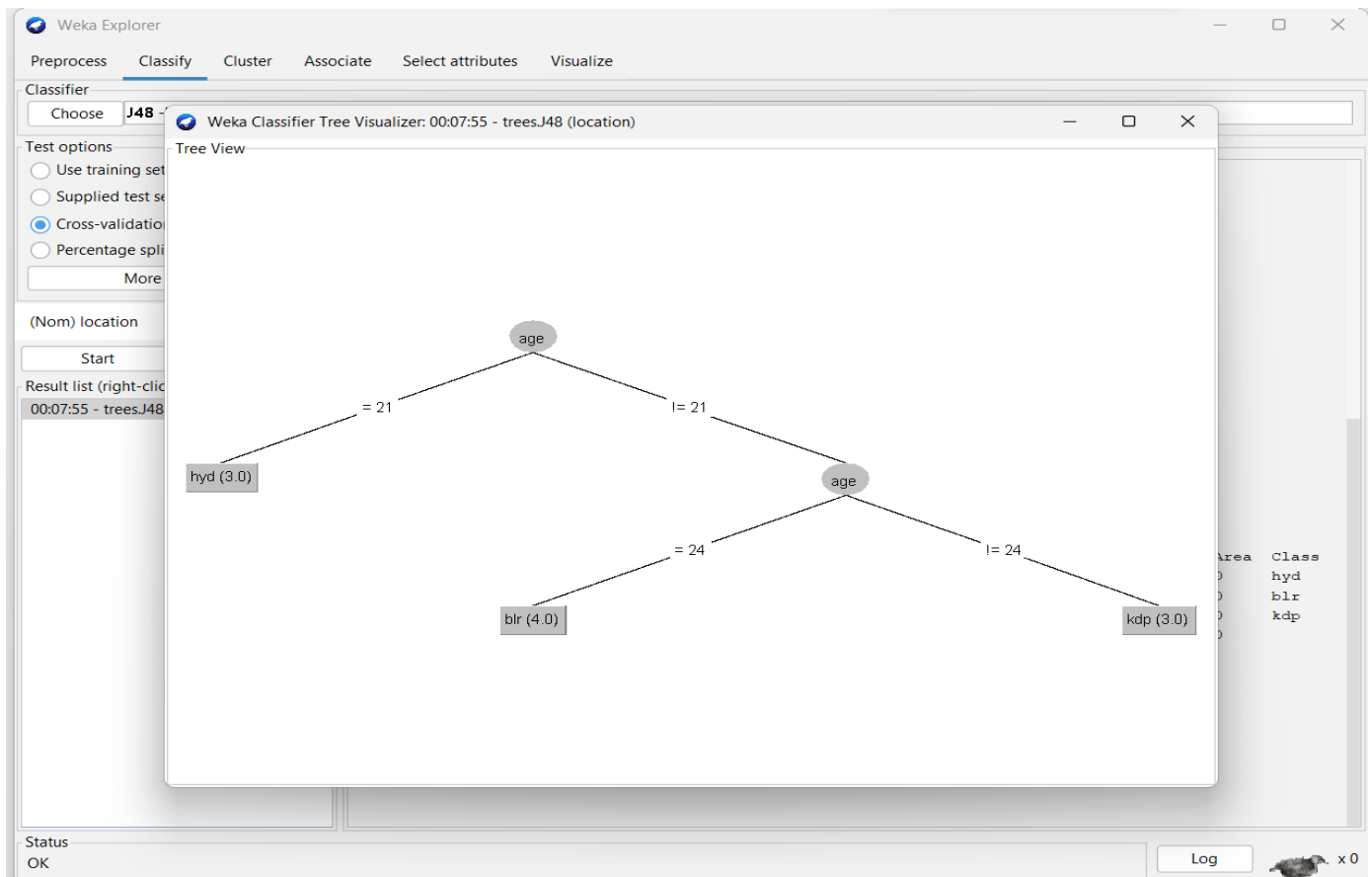
BinarySplit = True

Location:

BinarySplit = False

BinarySplit = True

CONCLUSION:

DISCUSSION AND VIVA VOCE:

- What is decision tree?
- Which algorithm is used in decision tree?
- What are the uses of decision tree?
- What is the advantage of decision tree?
- Is decision tree supervised or unsupervised and why?

REFERENCE:

- https://www.google.com/search?q=objective/in+apriori+algo&spell=1&sa=X&ved=2ahUKEwiOyqKGm5_-AhU1RmwGHfu6CUsQBSgAegQIBhAB&biw=1366&bih=600&dpr=1
- https://www.lucidchart.com/pages/decisiontree#:~:text=A%20decision%20tree%20is%20a,costs%2C%20probabilities%2C%20and%20benefits.

Data Mining – Concepts and Techniques, Jiawei Han & Micheline Kamber, Morgan Kauf-mann Publishers, Elsevier, 2nd Edition, 2006.

Observation book: (3)	Viva-Voce (3)	`Quality of Submission and timely Evaluation (4)
Total:		**Sign with date:**