OUTPUT

```
age = 25: poor(1.0)
age = 27: poor(2.0)
age = 28: poor(1.0)
age = 29: avg (2.0)
age = 30: avg (2.0)
age = 35: good (1.0)
```

age = 48 : good (2.0)

Number of Leaves : 7

Size of the tree : 8

Time taken to build model : 0 Seconds

== = Evaluation on training set == =

Time taken to test model on training data : 0 seconds

== = Summary = =

Correctly Classified Instances	11	100 %
Incorrectly classified Instances	0	0 %
Kappa statistic	1	
mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0 %	
Root relative squared error	0 %	
Total number of Instances	11	

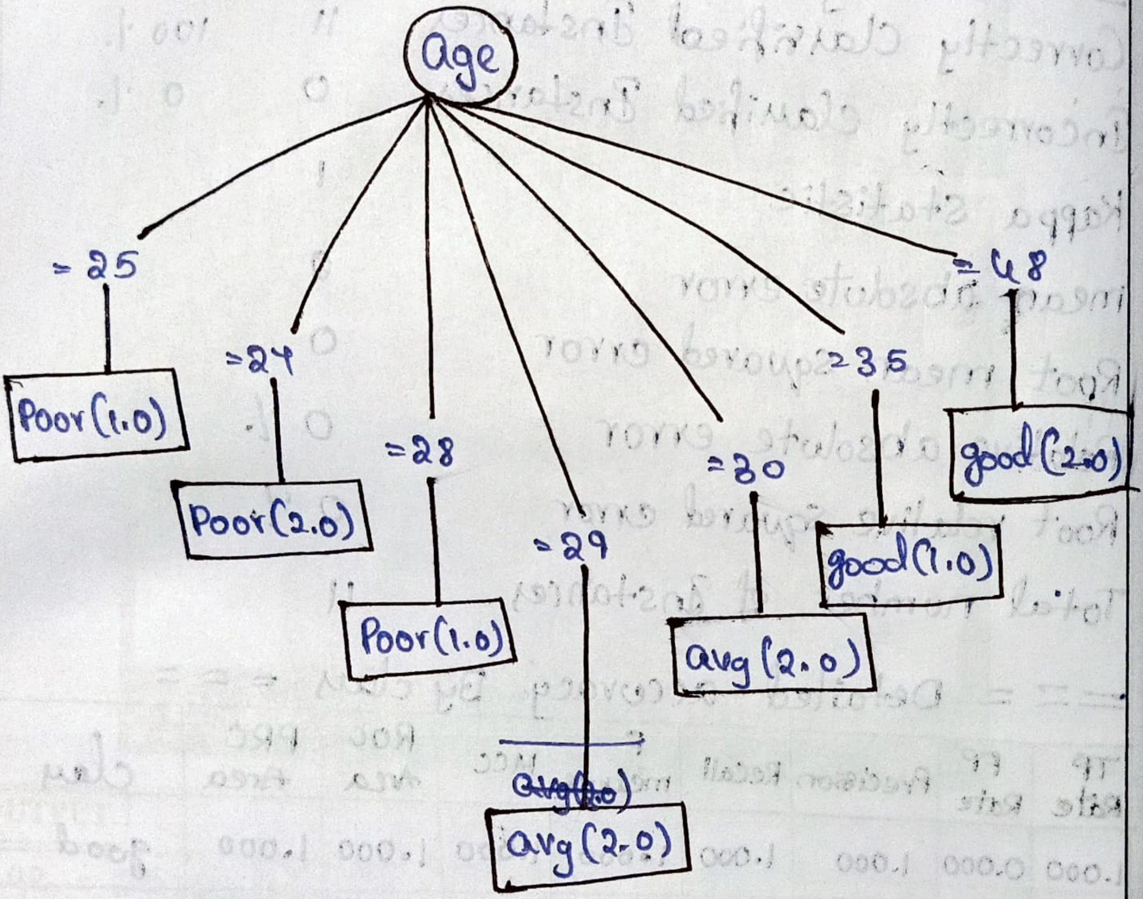
== = Detailed accuracy By class = = =

TP Rate	FP Rate	Precision	Recall	F measure	MCC	ROC Area	PRC Area	class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	good
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Avg
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	~~1.000~~ Poor
~~weighted avg~~	~~1.000~~ ~~0.000~~			~~1.000~~	~~1.000~~	~~1.000~~	~~1.000~~	
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

weighted avg

```
=== Confusion matrix ===
abc <-- classified as
300|a = good
040|b = avg
004|c = poor
```

after visualizing, Tree View



VIVA QUESTIONS

1. Define classification.

Ans. ● It is a data mining function that assigns items in a Collections to target Categories (or) classes

2. Define prediction.

Ans. Process of identifying the missing or unavailable numerical data for a new observation.

3. What are the classification techniques in data mining?

Ans. Classification, clustering, anamoly detection, regression, association rule learning & prediction

4. What are the advantages of different classification algorithms?

Ans. Efficient, not biased by outliers, no need for future scaling, Interpretability, works on non-linear problems

5. What is the kappa statistic? is a measure of how closely instances

Ans. classified by machine learning classifier matched the data labelled as ground truth, Controlling for the accuracy of a random classifier as measured by expected accuracy.

6. Which classification algorithm is best for prediction and analysis?

Ans. Random Forest

Week - 3

AIM:-

This experiment illustrates the use of j-48 classifier in weka The sample data set used in this experimentes "student" data available at arff format This document assumes that appropriate data pre processing has been Performed.

1) Load the data employee.arff into weka
2) Next, select the clasify tab & click on choose button
3) Then select "j48" classifier in trees Select use training set & to visualize tree right click & visualize tree.

```
@relation employee
@attribute age {25, 27, 28, 29, 30, 35, 48}
@attribute Salary {10k, 15k, 17k, 20k, 25k, 30k, 35k,
                                              32k}
@attribute performance {good, avg, poor}
@data
25, 10k, Poor
27, 15k, Poor
27, 17k, Poor
28, 17k, Poor
29, 20k, avg
30, 25k, avg
29, 25k, avg
30, 20k, avg
35, 32k, good
48, 35k, good
48, 32k, good
```

Output:-

```
age = 25 : poor(1.0)
age = 27 : poor(2.0)
age = 28 : poor(1.0)
age = 29 : avg (2.0)
age = 30 : avg(2.0)
age = 35 : good(1.0)
age = 48 : good (2.0)
Number of Leaves :7
Size of the tree : 8
Time taken to build model : 0 seconds
=== Evaluation on training set ===
Time taken to test model on training data : 0
                                        seconds

=== Summary ===
Correctly Classified Instances        11      100%
Incorrectly classified Instances       0        0%
Kappa statistic                        1
mean absolute error                    0
Root mean squared error                0
Relative absolute error                0%
Root relative squared error            0%
Total Number of Instances             11
```

=== Detailed accuracy. By class ===

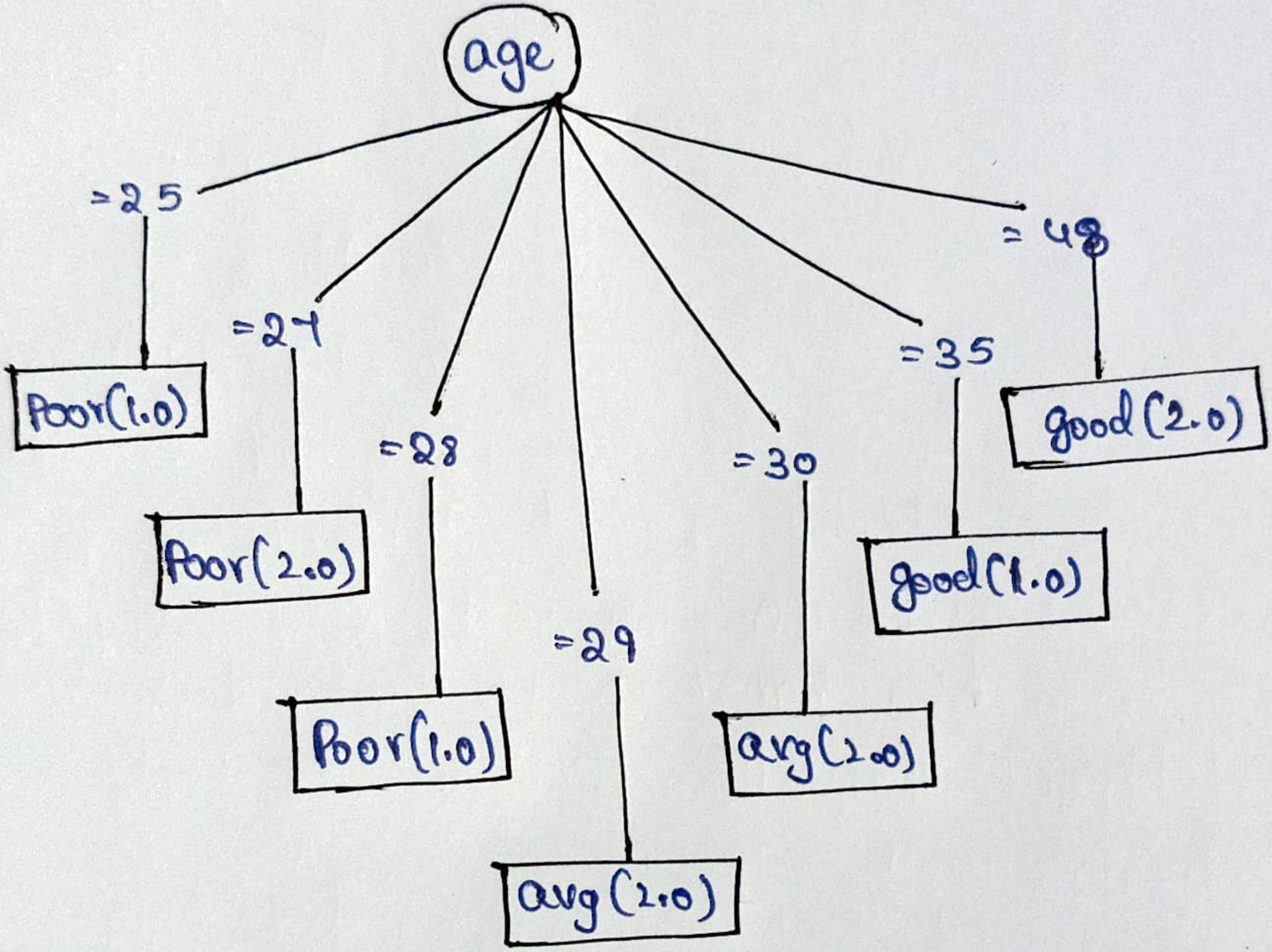
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	good
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	avg
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Poor
weighted avg	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion matrix ===

```
a b c  <-- classified as
3 0 0 | a = good
0 4 0 | b = avg
0 0 4 | c = poor
```

after visualizing tree:-

Tree View:


age
=25
=27
=28
=29
=30
=35
=48
Poor(1.0)
Poor(2.0)
Poor(1.0)
avg(2.0)
avg(2.0)
good(1.0)
good(2.0)