	Hope Foundation's **Finolex Academy of Management and Technology, Ratnagiri**		
	Department of Computer Science and Engineering (AIML)		
Subject name: Data Warehousing and Minig Lab			Subject Code: CSL503
Class	TE CSE	Semester –V (CBCGS)	Academic year: 2024-25
Name of Student			**QUIZ Score :**
Roll No		Experiment No.	03
Title: **Using open-source tools Implement Classifiers.**			

1. Lab objectives applicable:

LOB2: To introduce the concept of data Mining as an important tool for enterprise data management and as a cutting-edge technology for building competitive advantage.

2. Lab outcomes applicable:

LO3: Demonstrate an understanding of the importance of data mining.

LO6: Implement the appropriate data mining methods like classification, clustering or Frequent Pattern mining on large data sets.

3. Learning Objectives:

1. To perform classification by analyzing past data.

4. Practical applications of the assignment/experiment:

To analyze data and accordingly classify new/unseen data.

5. Prerequisites:

1. Java or Python or C programming language, mysql.

6. Minimum Hardware Requirements:

1. I series processor, RAM 4GB,

7. Software Requirements:

1. Weka 3.8

8. Quiz Questions: https://docs.google.com/forms/d/e/1FAIpQLScQhA-OT_avnpCs7iKm4RTVam7VCtcsg_OpsnM8tdfRuadeaA/viewform?usp=sf_link

9. Experiment/Assignment Evaluation:

Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A **or** any other relevant method.) Teacher should mention the other method used -		6
2	Lab Performance		2
3	Punctuality		2
Date of performance (DOP)		**Total marks obtained**	**10**

Signature of Faculty

11. Installation Steps / Performance Steps and Results –

Source code:

```
Classifier output
Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     bayesclassification
Instances:    14
Attributes:   5
              age
              income
              student
              credit_rating
              class
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

                Class
Attribute         yes     no
               (0.63) (0.38)
==============================
age
  youth            3.0    4.0
  middle           5.0    1.0
  senior           4.0    3.0
  [total]         12.0    8.0

income
  high             3.0    3.0
  medium           5.0    3.0
  low              4.0    2.0
  [total]         12.0    8.0

student
  yes              7.0    2.0
  no               4.0    5.0
  [total]         11.0    7.0

credit_rating
  fair             7.0    3.0
  excellent        4.0    4.0
  [total]         11.0    7.0
```

Viewer

Relation: bayesclassification

No.	1: age Nominal	2: income Nominal	3: student Nominal	4: credit_rating Nominal	5: class Nominal
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle	medium	no	excellent	yes
13	middle	high	yes	fair	yes
14	senior	medium	no	excellent	no

Add instance | Undo | OK | Cancel

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           8               57.1429 %
Incorrectly Classified Instances         6               42.8571 %
Kappa statistic                         -0.0244
Mean absolute error                      0.4374
Root mean squared error                  0.4916
Relative absolute error                 91.8631 %
Root relative squared error             99.6492 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.778    0.800    0.636      0.778    0.700      -0.026   0.578     0.697     yes
                 0.200    0.222    0.333      0.200    0.250      -0.026   0.578     0.557     no
Weighted Avg.    0.571    0.594    0.528      0.571    0.539      -0.026   0.578     0.647

=== Confusion Matrix ===

 a b   <-- classified as
 7 2 | a = yes
 4 1 | b = no
```

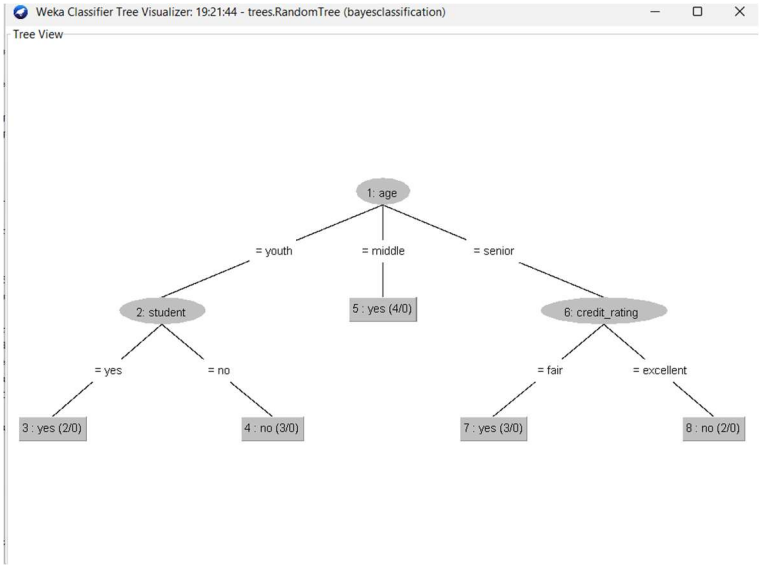
Viewer

Relation: bayesclassification

No.	1: age Nominal	2: income Nominal	3: student Nominal	4: credit_rating Nominal	5: class Nominal
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle	medium	no	excellent	yes
13	middle	high	yes	fair	yes
14	senior	medium	no	excellent	no

Add instance | Undo | OK | Cancel

Output: Screenshots



12. Learning Outcomes Achieved

1. Students are able to classify new tuple/data by performing analysis on historical data.

13. Conclusion:

1. Applications of the Studied Technique in Industry

- Predictive Maintenance: The classification technique can be applied in predictive maintenance within manufacturing industries to forecast equipment failures and schedule timely interventions, thereby reducing downtime and maintenance costs.

- Customer Segmentation: In the retail sector, classification techniques can be used to segment customers based on purchasing behavior, allowing for targeted marketing strategies and personalized customer experiences.

2. Engineering Relevance

- Efficiency Improvement: Implementing classification in engineering systems enhances decision-making processes, leading to improved operational efficiency and resource allocation.

- Quality Control: Classification techniques aid in quality control by identifying defects and anomalies in products, ensuring higher standards and reducing waste.

3. Skills Developed

- Data Analysis: Through the experiment, proficiency in data analysis was developed, enabling the interpretation and manipulation of large datasets to extract meaningful insights.

- Machine Learning Proficiency: The study honed skills in machine learning, particularly in understanding and applying various classification algorithms to solve real-world problems effectively.

14. References:

[1] https:// Paulraj Ponniah, "Data Warehousing: Fundamentals for IT Professional" , Wiley Publications

[2] Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann 3nd Edition.

[3] Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Person Education.

[4] Raghu Ramakrishnan and Johannes Gehrke, "Database Management Systems", 3rd Edition McGraw Hill.

[5] Elmasari and Navathe, "Fundamentals of Database Systems", Pearson Education.