Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	TE	**Semester:**	V
Course Code:	CSC504	**Course Name:**	Data WareHousing & mining

Name of Student:	Pranita Kumbhar
Roll No. :	70
Assignment No.:	03
Title of Assignment:	Classification
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	5
Demonstrated Knowledge	3	3
Legibility	2	2
Total	10	10

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge Legibility	3	2	1
Legibility	2	1	0

Checked by

Name of Faculty : Ms. Neha Raut

Signature :

Date :

Assignment 3:

Q.1] Explain the steps of Min-max normalization and apply them to transform an attribute value from an old range to a new range.

→

Min-Max normalization is a technique used in data preprocessing to rescale the range of features (attributes) into a specific range, usually [0,1] or [-1,1]. This ensures that no feature dominates others because of its scale.

- Steps of Min-Max Normalization:

1] Identify the old minimum (Old_min) and maximum (Old_max) values of the attribute.

2] Decide the new-range (New-min, New-max) where you want to scale the values.

3] Apply the formula:

$$V_{new} = \frac{(V_{old} - Old_{min})}{(Old_{max} - Old_{min})} \times (New_{max} - New_{min}) + New_{min}$$

where,

V_{old} = original value.

V_{new} = normalized value.

Example :
Suppose we have exam marks in the range [20, 80], and we want to normalize them into [0, 1].

- Old_min = 20,
- Old_max = 80,
- New_min = 0,
- New_max = 1
- Suppose Vold = 50

$$V_{new} = \frac{(50-20)}{(80-20)} \times (1-0) + 0$$

$$= \frac{30}{60}$$

$$V_{new} = \underline{\underline{0.5}}$$

The normalized value of 50 in [0, 1] range is 0.5.

Q.2] Explain the term True Positive, False Positive, True Negative, and False Negative in a confusion matrix. Apply them in the case of spam email detection.

→

A confusion matrix is used to evaluate the performance of a classification model. It compares the predicted values with actual values.

- Confusion matrix for Spam detection:

	Predicted Spam	Predicted Not Spam
Actual Spam	TP	FN
Actual Not Spam	FP	TN

Q.3] For the given dataset. Apply Naive - Bayes algorithm and predict the outcome for Car = (Red, Domestic, SUV).

Car	Type	Origin	Stolen
Red	Sports	Domestic	Y
Red	Sports	Domestic	N
Red	Sports	Domestic	Y
Yellow	Sports	Domestic	N
Yellow	Sports	imported	Y
Yellow	SUV	imported	N
Yellow	SUV	imported	Y
Yellow	SUV	Domestic	N
Red	SUV	imported	N
Red	Sports	imported	Y

→

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

$$X = [Red, Domestic, SUV].$$

- True Positive (TP):
The model correctly predicts a positive class.
- In spam detection:
An email is actually spam and the model predicts it as spam.

- False Positive (FP):
The model incorrectly predicts positive when it is negative.
- In spam detection:
An email is not spam (ham), but the model predicts it as spam. (Type 1 error).

- True Negative (TN):
The model correctly predicts the negative class.
- In spam detection:
An email is not spam and the model predicts it as not spam.

- False Negative (FN):
The model incorrectly predicts negative when it is positive.
- In spam detection:
An email is spam but the model predicts it as not spam. (Type II error).

i] $P(\text{Red} / \text{Yes}) = \dfrac{P(\text{Yes} / \text{Red}) \cdot P(\text{Red})}{P(\text{Yes})}$

$$= \frac{3/5 \cdot 5/10}{5/10}$$

$$= \underline{\underline{3/5}}$$

ii] $P(\text{Domestic} / \text{Yes}) = 2/5$.

iii] $P(\text{SUV} / \text{Yes}) = 1/5$.

$$P(X / \text{Yes}) = \frac{3}{5} \times \frac{2}{5} \times \frac{1}{5} = \frac{6}{125}$$

$$= \underline{\underline{0.024}}$$

i] $P(\text{Red} / \text{NO}) = \dfrac{P(\text{No} / \text{Red}) \cdot P(\text{Red})}{P(\text{NO})}$

$$= \frac{\frac{2}{5} \cdot \frac{5}{10}}{\frac{5}{10}}$$

$$= \underline{\underline{2/5}}$$

ii] $P(\text{Domestic} / \text{No}) = 3/5$

iii] $P(\text{SUV} / \text{NO}) = 2/5$.

$$P(X / \text{NO}) = \frac{2}{5} \times \frac{3}{5} \times \frac{2}{5} = \underline{\underline{0.072}} \quad \underline{\text{NO}}.$$

The predicted outcome for car = [Red, Domestic, SVV]

for Yes = 0.024

and No is 0.072.