BENAZIR BHUTTO SHAHEED UNIVERSITY LYARI

DEPARTMENT OF COMPUTER SCIENCE

# DATA MINING

Submitted to:

Ma’am Ambreen

Submitted by:

Hanifa Saleh Mohammad

CS 6th A

R. No: 672

Date: September,2021

Assignment i#3

**Ans-1: i**

In 3rd lecture we discussed about classification of data mining. Classification in data mining is a common technique that separates data points into different classes. It allows you to organize data sets of all sorts, including complex and large datasets as well as small and simple ones. It primarily involves using algorithms that you can easily modify to improve the data quality.

**Ans-2:**

**Confusion Matrix – Example**

• Imagine that you have a dataset that consists of 33 patterns that are

'Spam' (S) and 67 patterns that are 'Non-Spam' (NS).

• In the example 33 patterns that are 'Spam' (S), 27 were correctly

predicted as 'Spams' while 6 were incorrectly predicted as 'non-Spams'.

• On the other hand, out of the 67 patterns that are 'non-Spams', 57 are

correctly predicted as 'non-Spams' while 10 were incorrectly classified as

'Spams'.

* Accuracy = (TP+TN)/total = (27+57)/100 = 84%
* Misclassification Rate = (FP+FN)/total = (6+10)/100 = 16%
* True Positive Rate = TP/actual yes = 27/33 = 0.81
* False Positive Rate =FP/actual no = 10/67 = 0.15

**Ans-3**:

the classes of Transportation mode below consist of three groups of Bus, Car, and Train. In this case, we have 4 buses, 3 cars, and 3 trains (in

short we write as 4B, 3C, 3T). The total data is 10 rows.

– Prob(Bus) = 4/10 = 0.4

– Prob(Car) = 3/10 = 0.3

– Prob(Train) = 3/10 = 0.3

Entropy = ∑j -pjlog2pj

Entropy = - 0.4log2(0.4) - 0.3log2(0.3) - 0.3log2(0.3) =1.571

Gini Index = 1-∑j pj2

Gini Index = 1 – ((0.4)2 + (0.3)2+ (0.3)2) = 0.660