

SPPU-BE-COMP-CONTENT – KSKA Git

BUSINESS-INTELLIGENCE ASSIGNMENT - NO.-2

Q1.) Write a short note on Evaluation of Classification Models.
ANS. Evaluating classification Model is the process of using specific metrics and techniques to assess how accurately a model categorizes data into pre-defined classes.

0 Evaluation Metrics:-

1. Accuracy.

- It is the simplest evaluation Metric.
- It is defined as the number of correct prediction divided by the total number of prediction multiplied by 100.
- FORMULA:-

$$\text{Accuracy (A)} = \frac{\text{Total Number of Predictions}}{\text{Number of Current Predictions}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

- It is defined as the measure of correct positive predictions.

• FORMULA:-

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. Recall.

- It is defined as the measure of how many actual positives are captured.

FORMULA:- $\text{Recall} = \frac{TP}{TP + FN}$

4. F1-Score:-

It is defined as the Harmonic mean of Precision and Recall.

FORMULA:-

$$F1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Confusion Matrix:-

→ It is a table depicting True Positive, True Negative, False Positive and False Negative.

⇒

	Actually Positive	Actually Negative
--	-------------------	-------------------

Predicted Positive →	True (TP) Positives	False (FP) Positives
Predicted Negative →	False (FN) Negatives	True (TN) Negatives

Q2) Write a short Note on : Bayesian Method.

ANS. Bayesian Method:-

Bayesian Methods are probabilistic Approaches based on Baye's Theorem, used for prediction and decision making.

FORMULA:-

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

where,

- $P(A/B)$: Posterior Probability.
- $P(B/A)$: Likelihood Probability.
- $P(A)$: Prior Probability.
- $P(B)$: Evidence.

Advantages:-

1. Works well with small data.

2. Easy to implement.

Dis-Advantage:-

1. Assumption of Independence is often Unrealistic.

Q3.) Explain Clustering Methods in Detail.

Ans.

Clustering is an Unsupervised Machine Learning Technique used to group similar data points together.

o Types of Clustering:-

1. Partitioning Methods

- Example : K-Means.

- Divides data into k -clusters.

- Each point belongs to the Nearest Centroid.

STEPS:-

1. Choose ' k ' (No. of Clusters)

2. Assign points to the nearest centroid (Cluster)

3. Update Centroids.

4. Repeat Until Convergence.

2. Hierarchical Clustering.

. It builds a tree like structure. i.e. (Dendrogram)

TYPES:-

① Agglomerative. (Bottom Up Approach)

② Divisive. (Top-Down)

3. Density Based Clustering.

. Example : DBSCAN, Algorithm.

. Forms clusters based on Dense regions.

. Handles noise well.

Q4.) Explain Apriori Algorithm.

ANS.

• Apriori Algorithm is used for Association Rule Mining to find frequent time-stamps in datasets (like Market Basket Analysis)

• Key Concepts :-

Support : Frequency of Itemset.

Confidence : Likelihood of Rule.

• FORMULA :-

$$\text{Confidence } (A \rightarrow B) = \frac{\text{Support } (A \cup B)}{\text{Support } (A)}$$

STEPS :-

1. Generate candidate itemsets.
2. Calculate Support.
3. Prune Infrequent Itemsets.
4. Repeat Until No new Itemsets.
5. Generate Association Rule.

EXAMPLE :-

IF people buy Bread, then they tend to buy Butter.

Rule :-

Bread \rightarrow Butter.

Advantage :-

Simple and Easy to Understand.