

Name : Tejas Redkar

PRN : 1032210937

Panel-C, Batch-C2

Roll No : PC-44

DEC Lab Assignment -7

* Problem Statement

Your objective is to perform data classification technique & calculate the performance such as Accuracy, Precision, Recall & F1 Score. Build Classifier models using different techniques, analyze the confusion matrix & compare these models. Also apply cross validation while preparing the training & testing datasets.

* Objectives:

- To build the classifier models using different techniques
- To calculate the confusion matrix
- To calculate the performance of algorithms.

* Conclusion:

In conclusion, the application of classification techniques on a suitable dataset allows for the evaluation of performance using various metrics

such as Accuracy, precision, Recall & F1 score, these metrics provide insights into the effectiveness of the classification model & can guide further improvements or adjustments. By understanding & utilizing these performance metrics, machine learning practitioners can make informed decisions when solving classification problems.

* FAQ's

Q1) By giving appropriate examples, explain the difference between a binary classifier & a multiclass classifier

Ans 1) Binary classifier:

- Definition: It classifies input into two distinct categories or classes.
- Example: Spam or not spam email classification

2) Multiclass classifier:

- Definition: It classifies input into more than two classes or categories.
- Example: Image recognition classifying objects like cat, dog, & bird.

Q2) What is a Decision Tree Classifier? What are some advantage of decision trees?

Ans Decision-Tree Classifier

• **Definition:** A tree-like model that makes decisions based on feature values to classify instances.

* Advantages of Decision Trees:

- 1) **Interpretability:** Easy to understand & Interpret
- 2) **No Data Normalization:** No need for data normalization
- 3) **Handles Non-Linear Relationships:** Can capture non-linear relationships in data.
- 4) **Handles Mixed Data Types:** Can handle both numerical & categorical data.
- ~~5) **Handles Mixed Data Types:** Can handle both numerical & categorical data.~~
- 5) **Feature Importance:** Provides information about the importance of features.

Q3) How does a decision tree work?

- Ans
- 1) **Splitting:** Selects the best feature to split the data based on certain criteria (e.g. Gini impurity or information gain)
 - 2) **Recursive Process:** ~~Continuous~~ Continues the splitting process recursively for each branch until a stopping condition

is met.

3) Leaf Nodes: Each terminal node (leaf) represents a class or decision

4) Decision making: To classify a new instance, follows the path from the root to a leaf based on feature values.

Q4) What is the difference between a decision tree & a random forest tree?

Ans 1) Decision Tree: A single tree that makes decisions based on features

2) Random Forest: Ensemble of multiple decision trees, each trained on a random subset of data, & combines their predictions for improved accuracy & robustness.

