Ex : 3.

Study of the classifier with respect to statistical parameters

Aim

To study and compare the performance of various classifiers using statistical parameters such as accuracy, precision, recall, f1-score and confusion matrix.

Objective

TO understand how different machine learning classifiers performs on a given dataset

TO evaluate the classifier using key statistical matrics

TO compare and interpret the performance of each classifier for model selection

Pseudo code

1.) Import required libraries (Pandas, sklearn, etc...)

2.) Load the data set

3.) Preprocess the data (handling missing values, Encoding, scaling)

4.) Split the dataset into training and testing data

5.) Define a list of classifier - KNN, SVM, Decision tree

6.) For each classifier

a.) Train the model on the training set

b.) Predict using the test sets.

c.) Calculate statistical parameters

- Accuracy
- Precision
- Recall
- F1-score

7.) compare results in a tabular format

8.) Analyze and determine the best performing classifier based on context

i.) Accuracy

The Proportion of correctly Predicted samples out of the total samples

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negative}}{\text{Total samples.}}$$

ii) Precision

of all samples Predicted as Positive how many were actually Positive

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positive} + \text{False Positive}}$$

iii) Recall

shows how well the model detects actual Positives

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positive} + \text{False Negati}}$$

iv) F1-score

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

Inference :-

i) Here SVM Performance Best model with accuracy 96% along with balanced Precission, recall and f1 score
It Shows similar performance for both training and testing set is called generalization

ii) KNN → Is minimal overfitting
Training slightly higher than ~~testing~~ but still close

iii) Decision tree :- Training ≃ 100%, Testing ≃ 94% → clean overfitting (common in decision trees without Pruning)

Result

The breast cancer dataset has be used to train KNN, SVM and decision tree and the statistical Parameters has been compared and inference ,

Evaluation Matrix

Data set used :- breast-cancer dataset using Sklearn

Algorithm used :-

->KNN K-nearest neighbours

->SVM

-> Decision tree

KNN Matrics

Accuracy :- 9500'/.

	Precision	recall	f_1-Score	Support
0	0.93	0.93	0.93	43
1	0.96	0.96	0.96	71
			0.95	114
accuracy				
macro avg	0.94	0.94	0.94	114
weighted avg	0.95	0.95	0.95	114

SVM classifier Report: accuracy :- 96'/.

	Precision	recall	F_1-Score	Support
0	0.93	0.95	0.94	43
1	0.97	0.96	0.96	71
accuracy.			0.96	114
Marco avg	0.95	0.96	0.95	114
weighted avg	0.96	0.96	0.96	114

Decision Tree classification Report:

accuracy :- 0.9

	Precision	recall	F_1-Score	Support
0	0.93	0.93	0.93	43
1	0.96	0.96	0.96	71
accuracy			0.95	114
Macro avg	0.94	0.94	0.94	114
weighted avg.	0.95	0.95	0.95	114.