

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.

Objectives

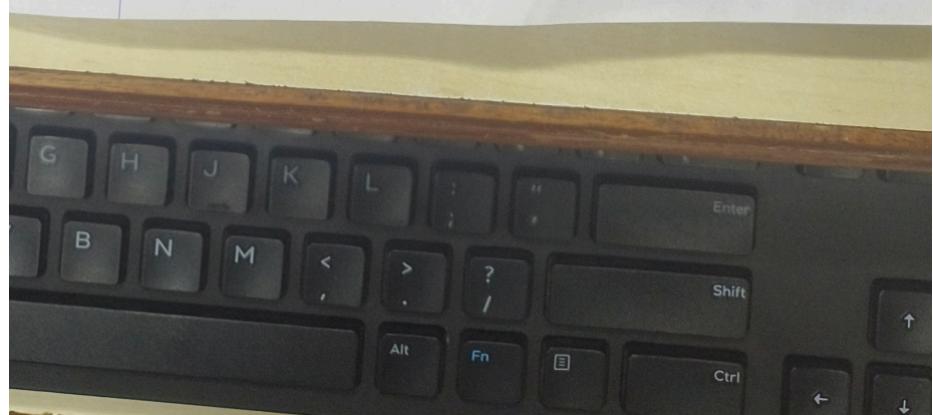
TO understand how different machine learning classifiers performs on a given dataset

TO evaluate the classifier using key statistical matrices

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



b.) For each classifier

- a.) Train the model on the training set

b.) Predict using the test set.

c.) calculate statistical parameters

- Accuracy

- Precision

- Recall

- F1-score

d.) compare results in a tabular format

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

i.) Accuracy

The proportion of correctly predicted samples out of the total samples.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

ii.) Precision

of all samples predicted as positive how many were actually positive

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

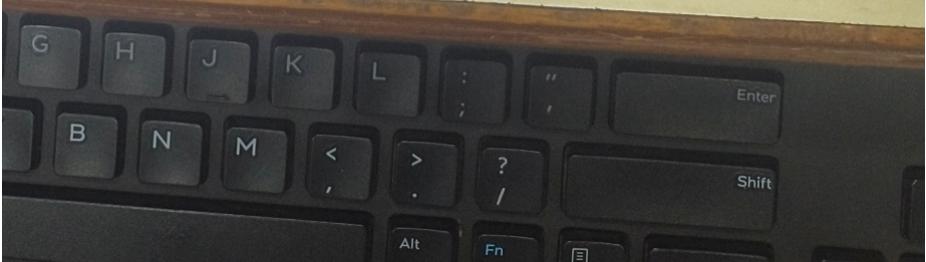
iii.) Recall

shows how well the model detects actual positives

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

iv.) F1-Score

$$F_1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

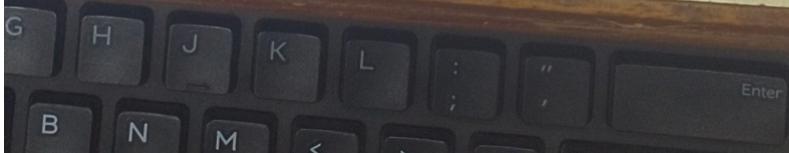


Inference :-

- i) Here SVM Performance Beat model with accuracy 96% along with balanced precision, 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 \approx 100% Testing \approx 94% \rightarrow clean overfitting (common in decision trees without pruning)

Result

The breast cancer dataset has been 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 neighbors

→ SVM

→ Decision Tree

KNN Matrix

Accuracy :- 93%

	Precision	recall	F1-Score	Support
0	0.93	0.93	0.93	43
1	0.96	0.96	0.96	71

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	F1-Score	Support
0	0.93	0.95	0.94	43
1	0.97	0.96	0.96	71

accuracy 0.96 114

macro 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	F1-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

