YCCE

Page No.

Date

PRACTICAL 6

Aim :- Performing Classification using sklearn for a given problem statement.

Problem Statement :- "Iris Flower Classification using KNN Algorithm"

The aim is to develop a ML model that can accurately classify Iris flowers into three distinct species : Setosa, Versicolor and Virginca based on four features : sepal's length & width, petal's length & width.

Using the Iris dataset from sklearn, K-Nearest Neighbour (KNN) algorithm will be implemented to classify flower species

Theory :- "Classification

- Classification is a process of categorizing data or objects into predefined classes or categories based on their features or attributes.

- Classification is part of supervised machine learning where you have input variable (x) and an output variable (Y) and use algorithm to map $Y = f(x)$ where labeled data is trained.

- The main objective of classification is to build a model to assign a label to a new observation based on its features.

☆ K-Nearest Neighbours Algorithm

- KNN is a simple, basic & instance-based learning algorithm. It belongs to supervised learning domain and finds instance application in pattern recognition, data mining.
- KNN algorithm is employed to tackle classification & regression problem. It is used for simplicity & handling of numerical & categorical data.

* Metrics of Distance Used in KNN Algorithm

1) Euclidean Distance :- It is cartesian distance between two points on a plane.

2) Manhatten Distance :- It is total distance traveled by object calculated by summing absolute distance between co-ordinates.

☆ Scikit-Learn

- It is a open source python library that provides a range of tools for various machine-learning tasks such as
  1. Classification
  2. Regression
  3. Clustering and many more.

YCCE
Page No.
Date

Need For Splitting the Data into Features & Target

- (i) Separating Input (x) & Output (Y)
- (ii) Training Model Properly
- (iii) Evaluation & Testing
- (iv) Modularity of Workflow

Need For Splitting the Data into Training & Testing Sets

- (i) Assessing Model Performance
- (ii) Preventing Overfitting & Pattern Memorization
- (iii) Improving Modal Tuning
- (iv) Providing an Unbiased Evaluation

Need For Scaling the Features

- (i) Equal Contribution of Features
- (ii) Improving Modal Convergence
- (iii) Impact on Distance-Based Algorithms like KNN
- (iv) Consistent Results Across Models
- (v) Avoiding Bias in Feature Importance