MINOR PROJECT REPORT

Submitted by : ABHIPSA JENA
Email : abhipsajena2k@gmail.com
Batch : JUNE 2021 - JULY 2021

PROBLEM STATEMENT:

To perform classification analysis on Iris dataset. Perform any classification algorithms and compare the accuracy.

PROJECT EXPLANATION:

<> A dataset called Iris dataset was provided to us, for analysis and classification.

<> The aim of the project is to use a classification algorithm to train the model and predict the species of Iris, using the sepal length, sepal width, petal length and petal width of the flower.

TOOLS USED:
Python
Jupyter Notebook

ALGORITHM:
The algorithm used for classification of the Iris data set is SVM - Support Vector Machine.

Support Vector Machine : The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

APPROACH EXPLANATION:

<> Iris dataset contains information about a flower called Iris, the information includes sepal length, sepal width, petal length, petal width and species of the Iris flower.

<> There are 3 distinct species of Iris present in the dataset, which means the dataset can be classified into 3 classes - Iris-setosa, Iris-versicolor and Iris-virginica.

<> Steps involved to achieve the final results include -

1. Exploration and analysis of the Iris dataset
2. Assignment of required x-values (sepal length, sepal width, petal length, petal width) and y-values(species)
3. Splitting the data into train and test data
4. Fit transforming x-values
5. Using SVC classifier to train the model with the training data
6. Using testing data on the model to predict the species of Iris
7. Comparing the predicted values with the actual values
8. Finally generating the confusion matrix and accuracy score.

CONFUSION MATRIX:

```
[[16,  0,  0],
 [ 0,  8,  0],
 [ 0,  1,  5]]
```

ACCURACY SCORE: 0.9666666666666667