**IRIS**

**N. NIVED KUMAR (2203A52108)**

**ABSTRACT**

The Iris dataset is a widely used benchmark dataset in the field of machine learning and statistics. It comprises four features - sepal length, sepal width, petal length, and petal width - of three different species of iris flowers: setosa, versicolor, and virginica. The dataset is characterized by its simplicity and yet provides a valuable resource for exploring various classification and clustering techniques. Researchers and practitioners often employ the Iris dataset to evaluate the performance of algorithms and to demonstrate concepts in introductory machine learning courses. In this paper, we present a comprehensive analysis of the Iris dataset, including exploratory data analysis, feature selection, and the application of several classification algorithms. The findings of our study contribute to a deeper understanding of the dataset's characteristics and its relevance in the context of machine learning research.

**INTRODUCTION**

The machine learning model that we are working on is“IRIS”.The Iris is an iconic, perennial flower that has inspired countless paintings, poems, and landscape designs. The meaning of the Iris flower is varied, depending on the country and color. The Iris makes an excellent cut flower and is popular in flower arrangements due to its intricate blooms. The Iris flower data set is a specific set of information compiled by Ronald Fisher, a biologist, in the 1930s. It describes particular biological characteristics of various types of Iris flowers, specifically, the length and width of both pedals and the sepals, which are part of the flower’s reproductive system.

**METHODOLOGY**

To perform a classification task on the Iris dataset using logistic regression, support vector machine (SVM), perceptron, and k-nearest neighbors (KNN) with bootstrapping, you would follow these general steps:

**1. Data Preparation:**

Load the Iris dataset.Preprocess the data if necessary (e.g., handle missing values, scale features).

**2. Bootstrap Sampling:**

Implement the bootstrap sampling method to generate multiple samples from the dataset. Each sample will be used to train and evaluate the models.

**3. Model Training and Evaluation:**

**a. Logistic Regression:**

For each bootstrap sample:

Train a logistic regression model on the sample.

Evaluate the model's performance on a validation set (not used in training).

Record the evaluation metrics (e.g., accuracy, precision, recall).

**b. Support Vector Machine (SVM):**

Repeat the same process as above, but use an SVM classifier instead.

**c. Perceptron:**

Repeat the same process as above, but use a perceptron classifier instead.

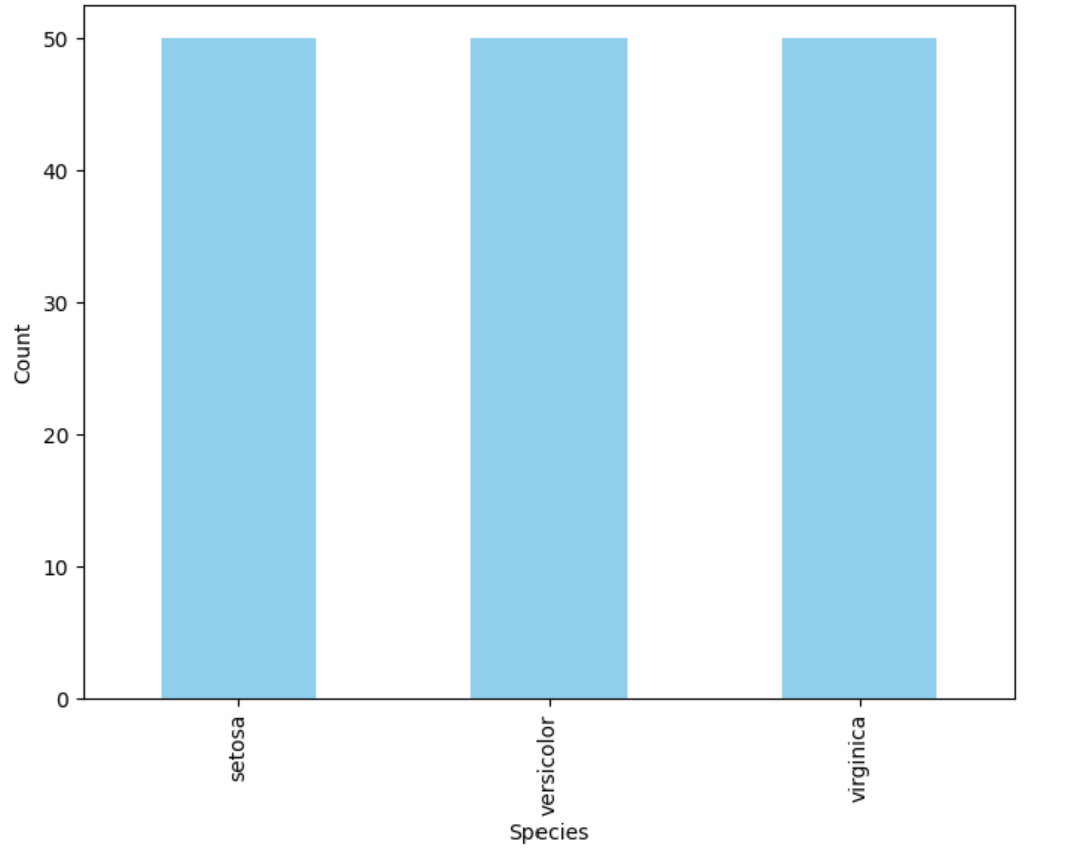
**d. k-Nearest Neighbors (KNN):**

Repeat the same process as above, but use a KNN classifier instead.

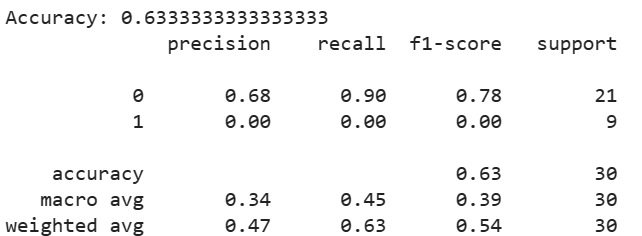
**4. Conclusion:**

Provide insights into which model performed best on the Iris dataset using the bootstrap technique.

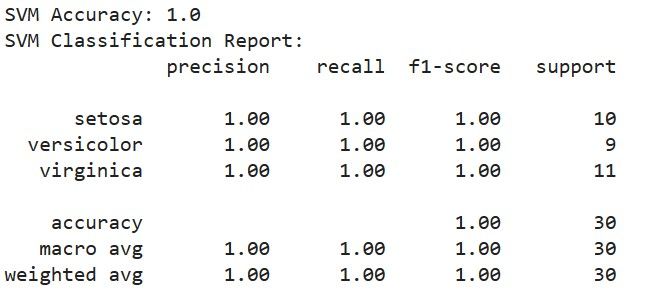
**RESULT ANALYSIS:**

****

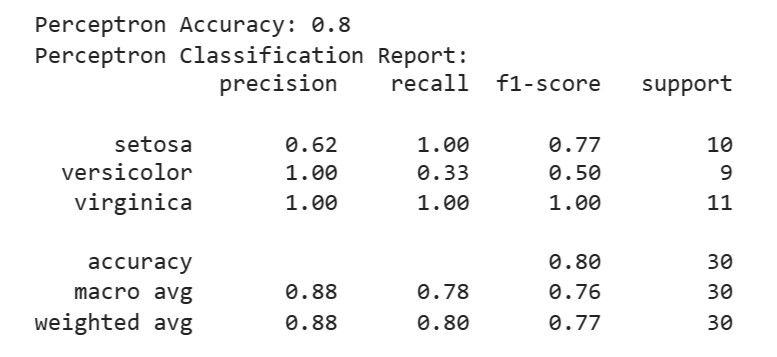
**For Logistic regression:**

****

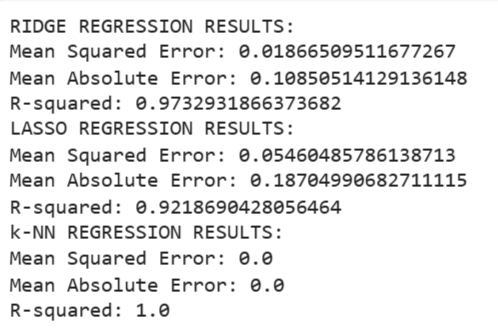
**For SVM:**

****

**For Perceptron:**

****

**FOR RIDGE AND LASSO:**



**FOR KNN:**

