Task Submission 2023 - June-P2 Batch OIB-SIP

Oasis Infobyte

Task 1 – Iris Flower Classification

By Sakshi Kharat

Problem Statement

- The objective is to develop a classification model that can correctly classify the different species of Iris Flowers using measurements of their Sepal Length, Petal Length, Sepal Width, and Petal Width.

- The dataset contains 150 examples of Iris Flowers from three different species that are Setosa, Versicolor, and Virginica.

Data Collection:

- The Iris dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems, and can also be found in the UCI Machine Learning Repository.

- It includes three iris species with 50 samples each as well as some properties of each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

Data Exploration and Visualization:

- An exploratory data analysis was performed to gain the details of the dataset prior to creating the classification models.

- This included analyzing the statistical summary of the features, visualizing the distributions of the features using Count Plot, Pair plots, Histograms, Box plots, and Heatmap, and exploring relationships between the features using scatter plots.

Algorithms used:

- **Logistic Regression:** Logistic regression is a supervised machine learning algorithm used for binary classification tasks. It is a linear model that predicts the probability of an input belonging to a particular class.

- **K-Nearest Neighbor Algorithm:** The k-Nearest Neighbors (KNN) algorithm is a supervised machine learning algorithm used for both classification and regression tasks. It is a non-parametric algorithm that makes predictions based on the similarity of data points in the feature space.

- **Decision Tree:** A decision tree is a supervised machine-learning algorithm that can be used for both classification and regression tasks. It creates a flowchart-like structure, where each internal node represents a feature, each branch represents a decision based on that feature, and each leaf node represents the outcome or prediction.

- **Random Forest Algorithm:** The Random Forest algorithm is an ensemble learning method that combines multiple decision trees to make predictions. It is a powerful algorithm widely used for classification tasks, including the classification of the iris flower dataset.

- **Support Vector Machine Algorithm:** Support Vector Machines (SVM) is a supervised machine learning algorithm that can be used for classification tasks, including the classification of the iris flower dataset. SVM is known for its ability to handle complex datasets and handle both linear and non-linear decision boundaries.
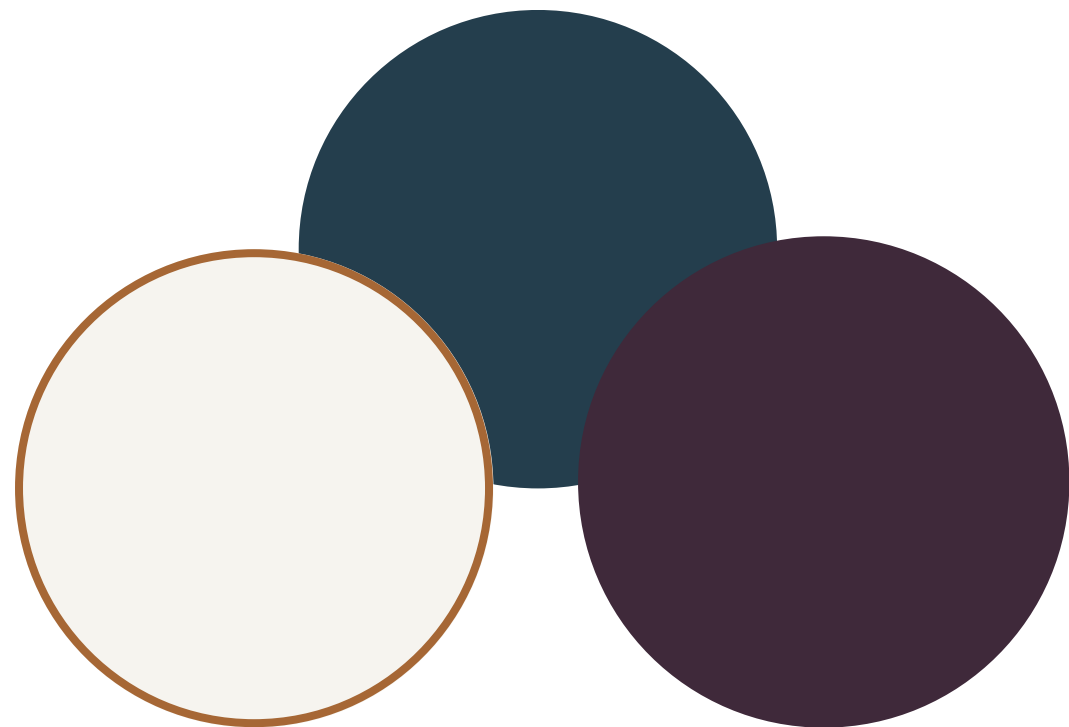
Accuracy Rate

Logistic Regression
Algorithm: 96.66

Accuracy of KNN
Algorithm: 100

Accuracy of Decision Tree
Algorithm: 96.66

Accuracy of Random Forest
Algorithm: 93.33

Accuracy of SVM
Algorithm: 96.66

Conclusion

- In conclusion, the iris flower dataset is a commonly used dataset in machine learning and is often used for classification tasks. I used various machine learning algorithms for the iris flower dataset to classify the different species accurately.

- I covered five popular algorithms: Logistic Regression, Decision Tree, Support Vector Machine (SVM), KNN, and Random Forest. Overall, the iris flower dataset serves as an excellent benchmark dataset for exploring and comparing different machine learning algorithms. The accuracy score of the above five models is very good and they can be used to predict the species of Iris Flower in five of the above models KNN Algorithm shows 100% accuracy.

THANKYOU ○

By : Sakshi Kharat