NAME:PRAJAKTA CHAVAN.
DATE:17/02/2024

# IRIS FLOWER CLASSIFICATION:

Iris flower has three species; setosa, versicolor, and virginica, which differs according to their measurements. Now assume that you have the measurements of the iris flowers according to their species, and here your task is to train a machine learning model that can learn from the measurements of the iris species and classify them.

Although the Scikit-learn library provides a dataset for iris flower classification, you can also download the same dataset from here for the task of iris flower classification with Machine Learning.

## description:

This project is part of the Data Science internship at **Oasis Infobyte**. It involves building a machine learning model for classifying Iris flowers into different species based on their sepal and petal measurements. The project includes data pre-processing, exploratory data analysis, model training, and deployment using treamlit for a user-friendly GUI.

## Dataset

The Iris dataset used in this project is sourced from Kaggle. The dataset contains measurements of sepal length, sepal width, petal length, petal width, and the corresponding species of the Iris flowers. The goal is to build a classification model that can accurately predict the species of an Iris flower based on its measurements.

Logistic regression is a process of modelling the probability of a discrete outcome given an input variable. Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors.

## Project Steps

Data Pre-processing: Load the Iris dataset and inspect its structure.

Remove unnecessary columns, such as an ID column, that do not contribute to the classification task. Perform any necessary data cleaning or handling of missing values.

Exploratory Data Analysis (EDA): Explore the dataset by visualizing the relationships between different features. Use plots, such as scatter plots, histograms, or box plots, to understand the

distributions and characteristics of each feature. Analyse any correlations or patterns in the data. Gain insights into the Iris flower species and their distinguishing features.

Feature Engineering: (Not needed in this project) If required, apply feature engineering techniques such as feature scaling, dimensionality reduction, or creating new features based on domain knowledge.

Data Splitting: Split the dataset into training and testing sets. The training set will be used to train the machine learning models, while the testing set will be used for evaluating their performance.

Model Training: Select multiple machine learning algorithms suitable for classification tasks, such as Logistic Regression, Decision Trees, and Random Forests. Train each model using the training data. Tune hype parameters, if necessary, using techniques like cross-validation or grid search.

Model Evaluation: Evaluate the trained models using the testing data. Calculate relevant evaluation metrics, such as accuracy, precision, recall, and F1-score, to assess the performance of each model. Compare the models and identify the best-performing one for further use.

Model Deployment: Extract the best-performing model along with any necessary pre-processing components, such as the label encoder and scaler. Use the Stream lit library to create a web-based GUI for the Iris flower classification model. Allow users to enter the sepal and petal measurements as input. Utilize the deployed model to predict the species of the Iris flower based on the user input.

## conclusion:

we have explored the process of building a machine learning model for iris flower classification. We have discussed the dataset, machine learning approach, pre-processing, training, and evaluation. By following these steps, we can create a model that accurately predicts the species of iris flowers based on their features. This technique can be applied to other classification problems as well, making it a valuable tool in the field of machine learning.