Name : Zeyad Mahmoud Abdelrahman Bassiouny

ID : 20201374047

Name : Abdelrahman Mohamed Essmat

ID : 20201374559

**Machine Learning Project**

In this project we used four machine learning models on the Iris dataset from Kaggle and evaluated each one.

The Iris dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems.

It includes three iris species with 150 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

Classes are :

* Iris Setosa
* Iris Versicolour
* Iris Virginica.

**Step 1** **:** Importing Libraries

We started by importing the packages needed such as pandas and sklearn.

**Step 2 :** Data Cleaning, Preparation And Splitting

We checked the data for null values and duplicate items. We also removed any unnecessary columns such as the ID column and split the data to targets (X) and features (Y).

**Step 3 :** Using The Machine Learning Models

Supervised Learning Models

**SVM Model :**

- We used a test size of 30% and a linear kernel with a degree of 3.

- The accuracy of the model is between 93% and 100%.

**KNN Model :**

- We used a test size of 30% and 3 neighbors.

- The accuracy of the model was between 93% and 100%.

**Naive Bayes Model :**

- In our last supervised learning model we also used a 30% test size and a random state of 42.

- The accuracy of the model was between 95% and 97%.

In conclusion, the supervised learning methods were almost identical to each other when working with the iris dataset, Although SVM was the most consistent among the three.

Unsupervised Learning Model

K-means Clustering Model :

In this clustering model we used K-value of 3 clusters, 300 iterations and a random state of 42.