1. R.A. Fisher's landmark 1936 study made use of the Iris dataset. It contains three iris species, each with 50 samples, as well as some information about each flower, such as the total number of instances (150) and the number of attributes (four). One flower species is linearly separable from the other two, but the other two are not.

The columns in this dataset are: Id, Sepal Length-Cm, Sepal Width-Cm, Petal Length-Cm, Petal Width-Cm and Species.

Perform the following:

a. Download the Iris dataset from kaggle.com.
b. Implement **Support Vector Machine** in python to classify the flower species using the following Kernel individually:
    a. Linear Kernel
    b. RBF Kernel
c. Use regularisation parameter=1
d. Compare the classification results for the input (Sepal Length in Cm=2.9, Sepal Width in Cm=4.1, Petal Length in Cm=2.5 and Petal Width in Cm = 08) using both the kernels.

e. Repeat the above four steps by reducing each iris species with 25 samples and comment on the classification results.

Attach the screen shots of Executable code and results.

2. Pima Indian diabetes data comes from the National Institute of Diabetes and Digestive and Kidney Diseases, which is part of the National Institute of Health. The goal of the dataset is to be able to diagnose whether or not a patient has diabetes based on certain diagnostic measurements that are in the dataset. There were a lot of rules that had to be followed when these people were chosen from a larger database. All of the patients here are women who are at least 21 years old and of Pima Indian heritage. The dataset includes 8 attributes and 1 Outcome and are: Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age and Outcome

Implement **Decision Tree classifier** in Python and perform the following steps:

1. Import the necessary Python Packages
2. Convert if any categorical attributes to Binary attributes
3. Display first 10 rows of the Pima dataset
4. Split the dataset into Training set:80% and Test dataset:20%
5. Train the model using Decision Tree Classifier.
6. Make the predations with test dataset
7. Find accuracy score, confusion matrix, precision, recall, f1-score, support and
8. Classify the input data (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age) = (5, 116, 74, 0, 0, 25.6, 0.201, 30)
9. Visualise the decision Tree.

Repeat the above steps for Training set:95% and Test dataset:5%

Attach the screen shots of Executable code and results.

3. Using the iris flower dataset, create a Random Forest model in Python to categorise the type of flower. It includes the sepal length, sepal breadth, petal length, petal width, and floral type. Setosa, versicolor, and virginia are the three species or classes. You may find the dataset in the Scikit-learn package or get it from the UCI Machine Learning Repository.

Implement **Random Forest Algorithm** in Python and perform the following steps:

1. Import the datasets
2. Print the labels and feature names
3. Separate the columns into dependent and independent variables
4. Split those variables into a training and test set (70% and 30% respectively)
5. Train the model on the training set.
6. Perform predictions on the test set.
7. Predict the accuracy of the model
8. Make a prediction for the input sample: sepal length = 4, sepal width = 3, petal length = 5, petal width = 1.5

Repeat the above steps by reducing each iris species to 25 samples and comment on the class prediction accuracy.

Attach the screen shots of Executable code and results.

4. Principal Component Analysis (PCA) is a way to address the issue "Curse of dimensionality in the machine learning domain" nothing but higher number of attributes in a dataset adversely affects the accuracy and training time of the machine learning model. PCA converts data from high dimensional space to low dimensional space by selecting the most important attributes that capture maximum information about the dataset.

Implement PCA in Python to achieve best classification by dimensionality reduction in breast cancer dataset which describe characteristics of the cell nuclei present in the image which has nearly 30 attributes with two classes benign and malignant.

Follow the steps:

1. Import all the necessary libraries
2. Load the dataset and display all the attributes on the screen
3. Construct a data frame.
4. Chose the Number of Principal Components = 3 and check dimensions of data.
5. Display the Eigen Vectors
6. Display Explained variance ratio

Attach the screen shots of Executable code and results.

5. Among Unsupervised Machine Learning Algorithms, **K means clustering** is one of the most popular. It is used to solve Classification Problems. K Means clustering divides the unlabelled data into different groups, called clusters, based on features and patterns that are similar to each other.

Implement K Means Clustering Algorithm for a randomly generated sequence.

Follow the steps:

1. Import the basic libraries
2. Generate 200 random numbers in 2 dimensional space
3. Train K Means algorithm by considering number of clusters equal to 3
4. Display the 3 clusters on 2D scattered plot.

5. Use Elbow Curve method to choose the number of centroids for improved clustering from the poor data we generated in step 1 by considering computational expense and knowledge gain from a dataset.
6. Mention the optimal number of clusters chosen and display improved clusters on 2D scattered plot

Attach the screen shots of Executable code and results.

6. Logistic regression is a statistical technique for modelling the probability of a specific class or occurrence.

> Social Network Ads is a categorical dataset describes information about a product being purchased through an advertisement on social media.
>
> Implement Logistic regression model in Python to predict whether the product is purchased or not by a person using any one of the three attributes given in the dataset.
>
> Implement the following steps:
>
> 1. Import the necessary libraries
> 2. Load the dataset
> 3. Consider any one highly related input attribute with the output variable and display the scatter plot
> 4. Use stochastic gradient decent method to train the model and use 300 epochs and initialize the weights=0 and learning rate=0.001, threshold value =0.5.
> 5. Plot the MSE for 300 epochs
> 6. Predict the MSE and accuracy of the trained model after 300 epochs
> 7. Validate the classification model for any 2 unseen values.

Attach the screen shots of Executable code and results.