MACHINE LEARNING
ASSIGNMENT - 4

NAME - Dharani Pulimamidi
700745377

Video Link -
https://drive.google.com/file/d/1d8DSwyXlExFS8xgjf2f7aK5Hoey0c2gJ/view?usp=share_link

1. **Pandas**

    1. Read the provided CSV file ‘data.csv’ .

    Df = pd.read_csv(‘C:\\Users\\dhara\\Downloads\\data (1).csv’) .

    2 . With describe() function from pandas module got the statistical description of data .

    3 . To check any null values in the data used isnull() function which results a Boolean value .

    4. Using agg() method we can apply certain operation on data . I applied aggregate functions on three columns like Pulse , Maxpluse and Calories .

    5. Using ‘ & ‘ operator we can filter the data based on conditions given .

    6. The values are greater than 500 and in pulse values are less than 100.

    7.Using copy() method we can copy the data from original file to another data file .

    8. For deleting the Maxpulse column , pop() method is used to remove the data.

    9. astype() method is used to convert from one data to another data .

    10. Using pandas created a scatter plot .

Titanic Dataset –

Using Python NumPy and Pandas libraries, I imported test and train data and combined them into a single dataset.

a. Do you think we should keep this feature?

The correlation results showed that males were strongly negatively correlated, and females were Strongly positively correlated with their survival. Males are inversely proportional, and females are directly proportional to their survival. So, we need this feature to analysis.

- I applied Naïve Bayes Algorithm on the preprocessed data using sklearn library. Python sklearn library contains many machine learning algorithms to analyze the data. In the given data, there are no labels present in the test data set to compare with our predicted data. So, we need to use the training data set to compare with our predicted data set using Naïve bayes algorithm. Using accuracy, we can compare with other machine learning algorithms to find which method is performing better on this data set.

Glass Dataset -

1. Implement Naïve Bayes method using scikit-learn library. Using read_csv method from pandas module I imported glass data set.

- Use train_test_split to create the training and testing part. sklearn module contains train_test_split method to split our data set into training and testing data sets. In this data set Type column can be used for labels. In this method, test_size defines how much proportion of data to be in the test data set. When we test_size value whole analysis results will change.

2. Evaluate the model on testing part using score and classification_report(y_ true, y_pred) In the given data, there are no missing values present in it. S

o, we can directly apply the machine learning algorithms on the data. sklearn module is imported to analyze the data using different algorithms. Classification_report and confusion_matrix methods to result the summary of the predictions made using the specific algorithm. These summaries can be used to compare with another algorithms to define which algorithm is better. The Accuracy by NavieBayes is 0.37209302325581395 and the accuracy by SVM is 0.5116279069767442. So SVM gives the better accuarcy.