**ML ASSIGNMENT 2**

Sai Susmitha Devi Komatlapalli

700742098

[**https://github.com/SusmitharajuK/ML\_Assingment\_2**](https://github.com/SusmitharajuK/ML_Assingment_2)

<https://drive.google.com/drive/folders/1nbAkZVnInPbXPbidGqt6_7Wm6YFJ2AwN?usp=sharing>

**1. Pandas**

1. Read the provided CSV file ‘data.csv’. <https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>

2. Show the basic statistical description about the data.

A screenshot of a computer

Description automatically generated

3. Check if the data has null values. a. Replace the null values with the mean

4. Select at least two columns and aggregate the data using: min, max, count, mean.

A screenshot of a computer

Description automatically generated with medium confidence

5. Filter the dataframe to select the rows with calories values between 500 and1000.

6. Filter the dataframe to select the rows with calories values > 500 and pulse.

A screenshot of a computer

Description automatically generated with medium confidence

7. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.

8. Delete the “Maxpulse” column from the main df dataframe

A screenshot of a computer

Description automatically generated

9. Convert the datatype of Calories column to int datatype.

10. Using pandas create a scatter plot for the two columns (Duration and Calories).

A screen shot of a computer

Description automatically generated with medium confidence

**2. Scikit-learn**

1. Implement Naïve Bayes method using scikit-learnlibrary.

a. Use the glass dataset available in Link also provided in your assignment.

b. Use train\_test\_split to create training and testing part.

2. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)

1. Implement linear SVM method using scikit library

a. Use the glass dataset available in Link also provided in your assignment.

b. Use train\_test\_split to create training and testing part.

2. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)

Do at least two visualizations to describe or show correlations in the Glass Dataset.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer screen

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Algorithm with better accuracy and justification:

We got better accuracy for Naïve Bayes method which is 0.8372093023255814. Naive Bayes analysis works well with probabilistic concepts whereas Linear SVM works better with linear regression logics. But to perform more accurately SVM requires large amounts of data to train and test the data. So, due to the amount of data Naive Bayes algorithm gives better accuracy compared to Linear SVM.