Machine Learning (Assignment 2)

Krishna Vamsi Koppula

700745021

GitHub Link: <https://github.com/Krishnavamsikoppula/Assignment-2-ML-Summer>

Video Link: <https://drive.google.com/file/d/1MaGHcxUD1UKF-tbTcTMpMgz-CgYmBUAu/view?usp=sharing>

**1. Pandas**

1. Read the CSV file ‘data.csv’.

A screenshot of a computer

Description automatically generated with medium confidence

2. Show the basic statistical description about the data.

A screenshot of a computer program

Description automatically generated with medium confidence

3. Check if the data has null values.

a. Replace the null values with the mean

A screenshot of a computer program

Description automatically generated with medium confidence

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 data frame to select the rows with calories values between 500 and1000.

A picture containing text, font, line, screenshot

Description automatically generated

6. Filter the data frame to select the rows with calories values > 500 and pulse <100.

A screenshot of a computer code

Description automatically generated with low confidence

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

A screen shot of a computer

Description automatically generated with low confidence

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

A screenshot of a computer

Description automatically generated with medium confidence

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

A picture containing text, screenshot, font, line

Description automatically generated

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

A screen shot of a graph

Description automatically generated with medium confidence

2. Scikit-learn

1. Implement Naïve Bayes method using scikit-learn library.

a. Use the glass dataset available in 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)

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

1. Implement linear SVM method using scikit library

a. Use the glass dataset available in 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)

A screenshot of a computer

Description automatically generated with medium confidence

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 graph

Description automatically generated with low confidence

Which algorithm you got better accuracy? Can you justify why?

Better accuracy is for Naive Bayes method 0.24615384615384617. Naïve Bayes analysis works well with probabilistic concepts where as linear SVM works better with linear regression values. So, Naïve Bayes algorithm gives better accuracy compared to linear SVM.