Summer 2023: ML 5710

(Assignment 2)  
**GitHub Link**: https://github.com/Rajesh-Adepu/assignment2.git

**Adepu Rajesh 700741537**

# Programming elements:

Classification

# Pandas

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

<https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>

A screenshot of a computer

Description automatically generated

* 1. Show the basic statistical description about the data.

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

* 1. Filter the dataframe to select the rows with calories values > 500 and pulse <100

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

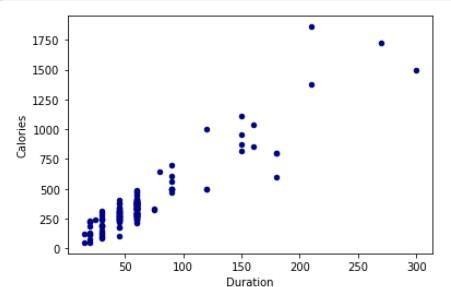
Description automatically generated

* 1. Delete the “Maxpulse” column from the main df dataframe

A screenshot of a computer

Description automatically generated

* 1. Convert the datatype of Calories column to int datatype.
  2. Using pandas create a scatter plot for the two columns (Duration and Calories).
     1. Example:



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

# Scikit-learn

* 1. Implement Naïve Bayes method using scikit-learn library.
     1. Use the glass dataset available in [Link](https://umkc.box.com/s/ea6wn1cidukan67t02j60nmp1ljln3kd) also provided in your assignment.
     2. 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
   1. Use the glass dataset available in [Link](https://umkc.box.com/s/ea6wn1cidukan67t02j60nmp1ljln3kd) also provided in your assignment.
   2. 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. Which algorithm you got better accuracy? Can you justify why?

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated