Library imports

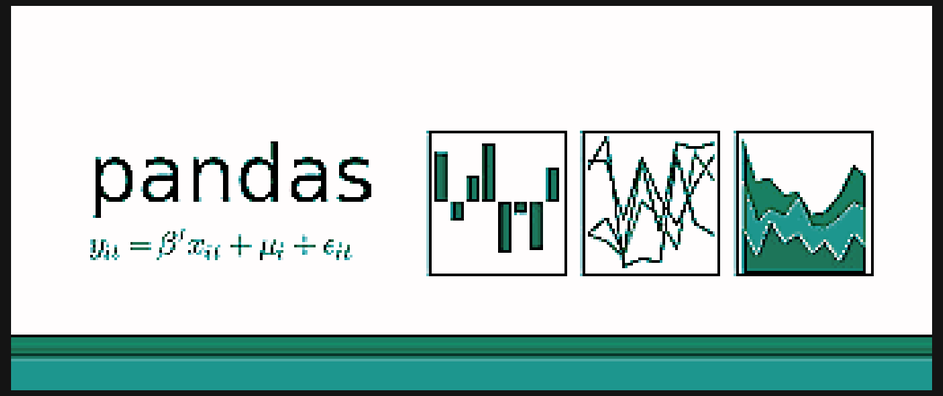
BHAVANA TADIGADAPA
ANAND PATEL

Some of Libraries needed to start the implementation Logistic Regression

- IMPORT PANDAS AS PD
- IMPORT NUMPY AS NP
- IMPORT MATPLOTLIB.PYPLOT AS PLT
- IMPORT SEABORN AS SNS

Pandas

IMPORT PANDAS AS PD



IT IS A PACKAGE WHICH HAS NUMEROUS TOOLS FOR DATA ANALYSIS AND IT CONTAINS MANY DATA STRUCTURES. SO HERE IT CAN BE USED FOR DATA ANALYSIS AND MANIPULATION.

FOR REFERENCE OF PANDAS:HTTPS://PANDAS.PYDATA.ORG/PANDAS-DOCS/STABLE/REFERENCE/INDEX.HTML

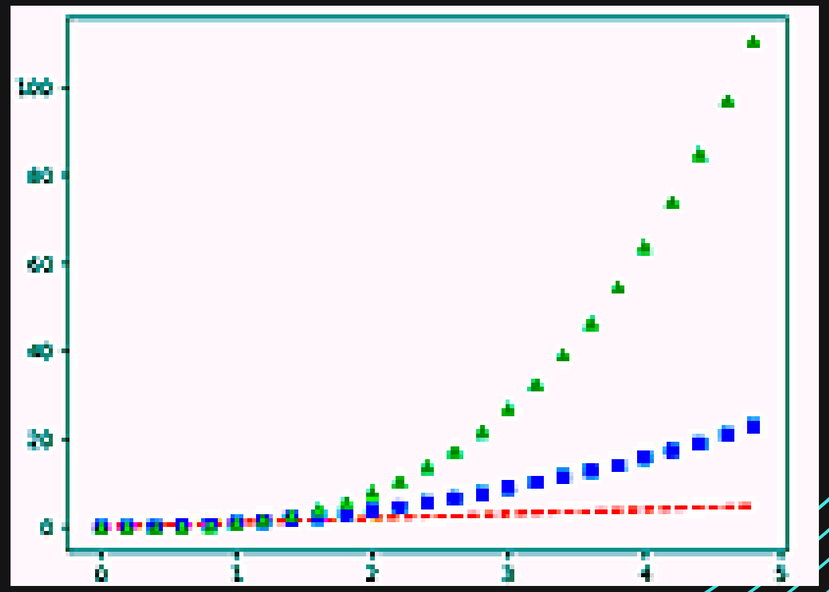
NumPy

IMPORT NUMPY AS NP

IT IS USEFUL FOR PERFORMING MATHEMATICAL AND LOGICAL OPERATIONS ON ARRAYS. WE CAN CREATE NUMPY ARRAYS, ACCESS VALUES AND MANIPULATE ARRAYS.

FOR REFERENCE OF NUMPY:
HTTPS://WWW.W3SCHOOLS.COM/PYTHON/NUMPY_INTRO.ASP

Matplotlib.pyplot



- **Matplotlib is a plotting library for Python.**
- It provides an object oriented API(Application programming interface) for embedding plots into applications.
- Pyplot is a collection of command style functions that make matplotlib work like MATLAB.
- .Each pyplot function makes some change to the figure.
- **For example three sets of data are being plotted each in a different way as shown**

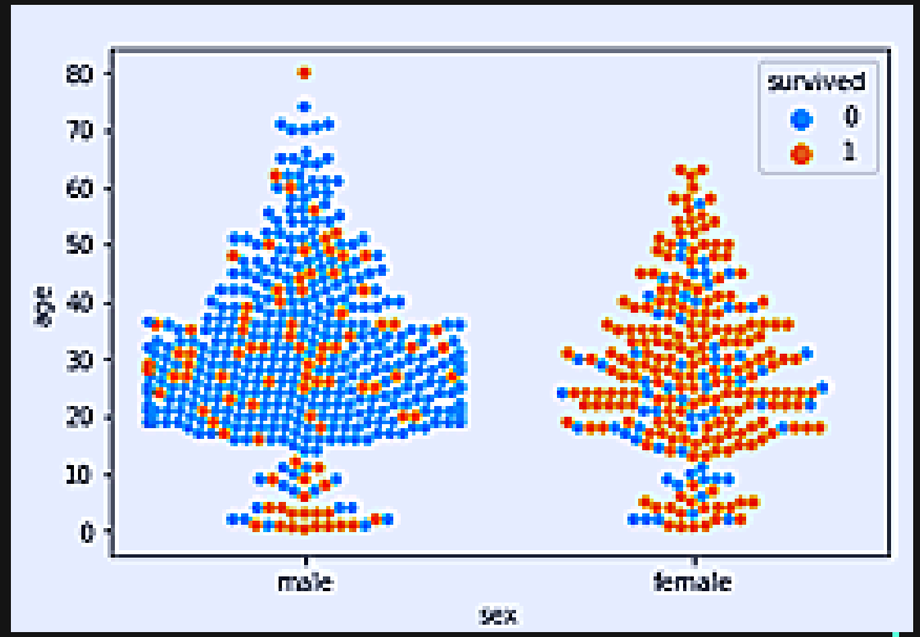
FOR FURTHER REFERENCE OF PYLOT:
HTTPS://MATPLOTLIB.ORG/3.1.1/TUTORIALS/INTRODUCTORY/PYPLOT.HTML

Seaborn

- It was developed based on matplotlib library. It is used to create more attractive and informative statistical graphics while representing a plot.
- It uses matplotlib underneath to plot graphs. It shows graph in a more visualized way.

For example a plot for survived persons in the titanic boat is shown.

FOR FURTHER REFERENCE OF SEA BORN:
HTTPS://STACKABUSE.COM/SEABORN-LIBRARY-FOR-DATA-VISUALIZATION-IN-PYTHON-PART-1/L



FROM THE OUTPUT, IT IS EVIDENT THAT THE RATIO OF SURVIVING MALES IS LESS THAN THE RATIO OF SURVIVING FEMALES.

- In the csv file of the data set we have the data of the columns:

- Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin,BMI,DiabetesPedigreeFunction,Age,Outcome

- col_names=['pregnant','glucose', 'bp','skin', 'insulin', 'bmi','pedigree', 'age', 'label'] is to label each column of csv file as csv has data with no header.

- **Reads the data from the csv file and stores in diabetes_data. Header=None is used here because by default the first row is taken as header from the datafile. Names=col_names is to give header names for each column.**

Selecting Feature and Splitting Data

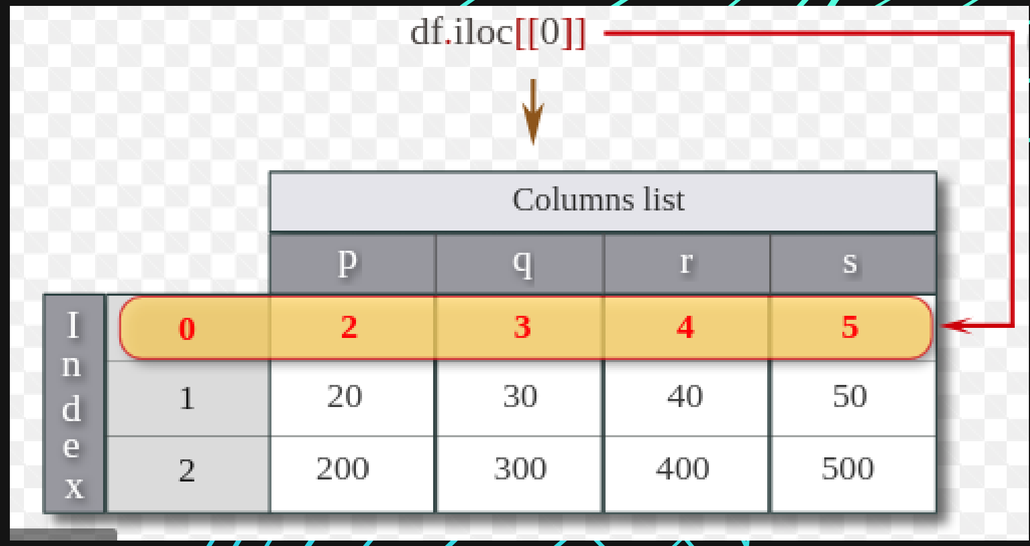
Selecting Feature and Splitting Data?

- ELIMINATING FIRST ROW AND DISPLAYING DATAFRAME
- SELECTING FEATURE
- SPLITTING OF DATASET IN FEATURE AND TARGET VARIABLE
- DIVIDING THE DATASET INTO A TRAINING SET AND A TEST SET

WHAT IS ILOC?

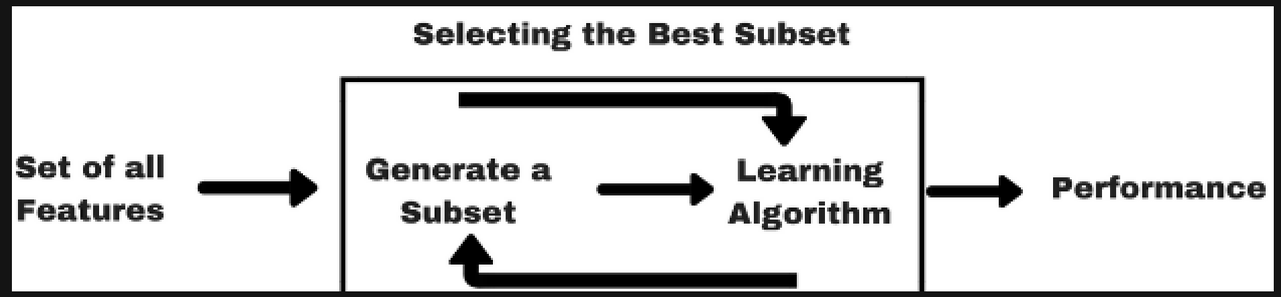
- Pandas provide a unique method to retrieve rows from a Data frame.
- Rows can be extracted using an imaginary index position which isn't visible in the data frame.

Syntax: pandas.DataFrame.iloc[]



Reference Link:https://www.geeksforgeeks.org/python-extracting-rows-using-pandas-iloc/

WHY SELECTING FEATURE IS IMPORTANT?



- it enables the machine learning algorithm to train faster.
- It reduces the complexity of a model and makes it easier to interpret.
- It improves the accuracy of a model if the right subset is chosen.
- It reduces overfitting.

What Sklearn and Model_selection are?

- Sklearn is a Python library that offers various features for data processing that can be used for classification, clustering, and model selection.

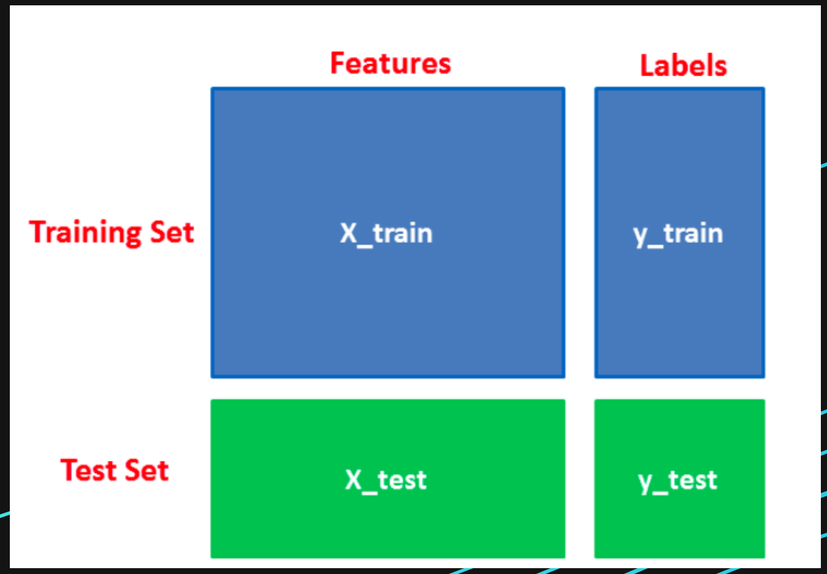
- method for setting a blueprint to analyze data and then using it to measure new data.

- Selecting a proper model allows you to generate accurate results when making a prediction.

REFERENCE LINK: HTTPS://WWW.BITDEGREE.ORG/LEARN/TRAIN-TEST-SPLIT

What is train_test_split?

- Is a function in Sklearn model selection for splitting data arrays into two subsets: for training data and for testing data.
- The train-test split is a technique for evaluating the performance of a machine learning algorithm.
- It can be used for classification or regression problems and can be used for any supervised learning algorithm.



THANK YOU