\*\*\*\*REGRESSION\*\*\*\*

Extract data from diabetes dataset of sklearn library and show it as a complete dataset with proper column header and target column

Find sum of null values in each column of above dataset

Remove the duplicate rows

Now divide the dataset into independent and dependent columns

Create correlation matrix for independent columns

Create correlation heatmap

Remove s1 from independent columns and then show the remaining independent columns

Split independent and dependent columns into train and test

Normalise independent columns using StandardScalar

Now select following models and check the root means square error and adjusted r square with respect to all

LinearRegression, Ridge, Lasso, ElasticNet

Select the model which has lowest root mean square error and highest adjusted r square

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*CLASSIFICATION\*\*\*\*

Extract data from load\_breast\_cancer dataset of sklearn library and show it as a complete dataset with proper column header and target column

Now divide the dataset into independent and dependent columns

Split independent and dependent columns into train and test

Normalise independent columns using StandardScalar

Now select following models and check the accuracy, precision, recall and f1 score with respect to all

Logistic Regression, Naive Bayes, Decision Tree, Random Forest and Support Vector Machine

Select the model which has highest accuracy, precision, recall and f1 score