**Heart Attack Possibility Prediction**

**Milestone: Model selection and implementation**

Group 18

Student 1 Karan Paresh

Student 2 Tianyu Yang

857-294-7660 (Tel of Karan Paresh)

617-368-0332 (Tel of Tianyu Yang)

[samani.k@northeastern.edu](mailto:samani.k@northeastern.edu)

[yang.tianyu@northeastern.edu](mailto:yang.tianyu@northeastern.edu)

**Percentage of Effort Contributed by Student 1: 50%**

**Percentage of Effort Contributed by Student2: 50%**

**Signature of Student 1: Karan Paresh**

**Signature of Student 2: Tianyu Yang**

**Submission Date: 10/4/202**

**Model selection and implementation**

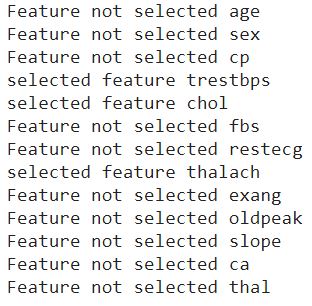
**Feature selection: -**

1. Lasso L1 penalty feature selection: -

The L1 penalty feature selection is used to determine the features that are needed that is the

once that influence the target variable the most are selected.The features whose coefficients

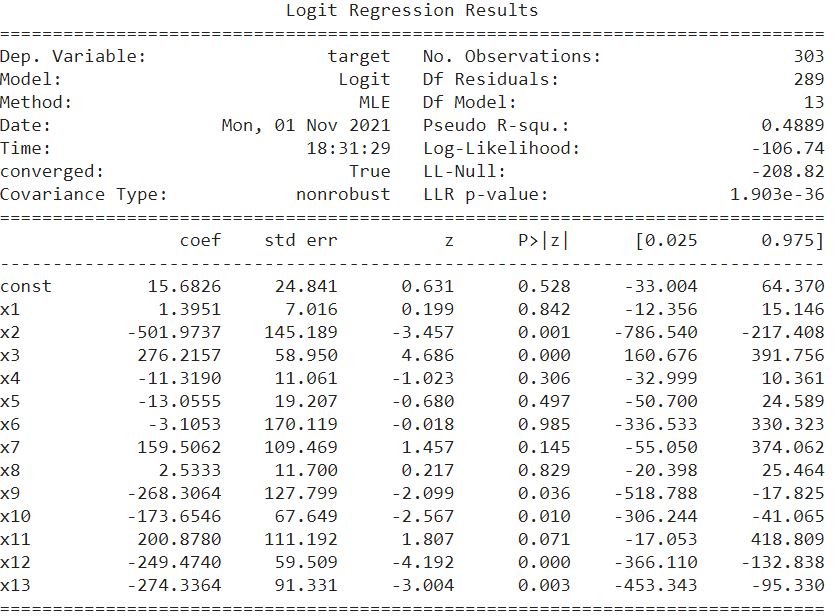
shrank to zero are removed and the once which didn't are selected.

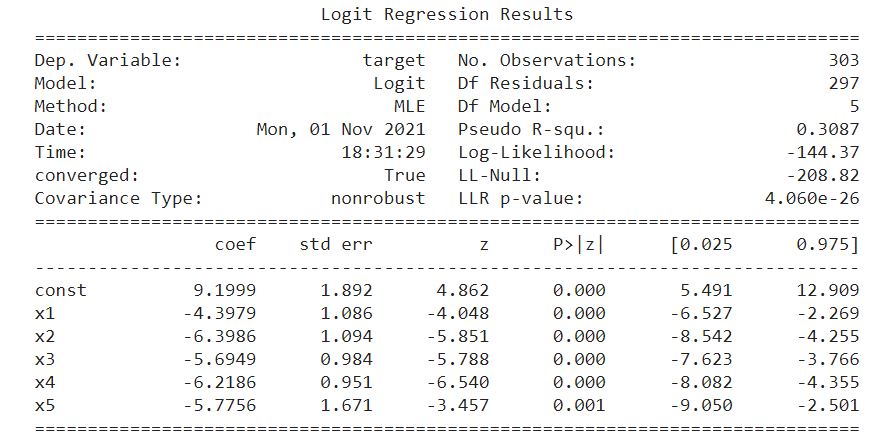


The features which are selcted by the L1 penalty are "trestbps", "chol" and "thalach".

2. Backward feature selection: -

We use the other method known as backward feature selection, the selection of a feature her depends on its p-value. In this method at first all the features are selected and the p-value for every feature is noted, if the p-value of any variable greater than that of "0.05" that feature removed from the group and features with p-value lesser than "0.05" are kept and tested again and we continue until we reach a point where all the feature has p-value lesser than "0.05", once we achieve this we stop and select those features.



Final result,

The Selected features are 'sex', 'exang', 'oldpeak', 'ca' and 'thal'. There is a difference when we use a different method for feature selection. There is no particular method it all depends on the model and the data, so we need to try different methods. We can also use "ridge regression", "forward selection" and etc.

**Model selection: -**

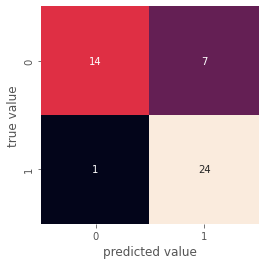
1. Gaussian Naïve Bayes: -

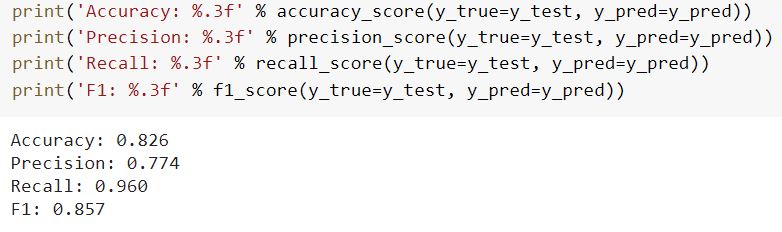
We use the gaussian naive bayes model to classify the class whether the patient will suffer from "Heart attack" or not. We use this model because every feature in this data is independent of each other and its relation to the target variable is also independent. The model is able to predict the results with 82.6% accuracy which is relatively decent.

Evaluation: -

Confusion matrix and accuracy score: -

Target classes = 1- High chances Heart attack, 0- Less chances of heart attack.



Accuracy Scores: -

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Precision = (TP)/(TP+FP)

Recall = (TP)/(TP+FN)

F1 = (2\*Recall\*Precison)/(Recall+Precision).

The model provides us with a decent accuracy score in distinguishing the class of a patient which helps in determining the chances of heart attack without going through the medical test by just using the values of the data of a patient.