

SC1015-Mini Project

Heart Attack Analysis and Computing Mahi Pandey & Yalamanchili Sanjana Team 7:FCE3

TABLE OF CONTENTS

01 PRACTICAL MOTIVATION

2 EXPLORATORY ANALYSIS

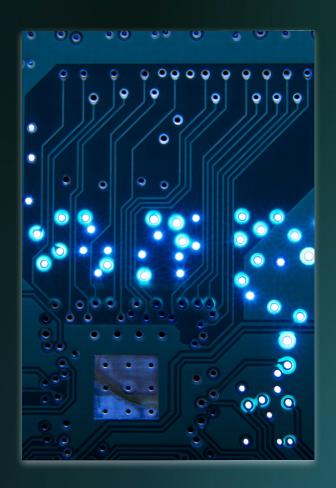
03 LOGISTIC REGRESSION

04 RANDOM FOREST

05 GRADIENT BOOSTING

06 INSIGHTS

PRACTICAL MOTIVATION





More than half a billion people around the world continue to be affected by cardiovascular diseases, which accounted for 20.5 million deaths in 2021¹ – close to a third of all deaths globally and an overall increase on the estimated 121 million CVD deaths.



PROBLEM DEFINITION

- Exploratory analysis of variables and Feature selection to statistically rank the features contributing to the risk of a heart attack.
- Computing the accuracies of two machine learning models and comparing on the basis of 13,8 and 3 variables.

INTRODUCING OUR DATASET

age Age of the Patient

trtbps Resting blood pressure (in mm Hg)

cho Cholesterol in mg/dl fetched via BMl sensor

thalachh Maximum heart rate achieved

Caa Number of Major Vessels Coloured by Fluoroscopy

INTRODUCING OUR DATASET

Sex Sex of the Patient

CP Chest Pain Type

fbs Fasting Blood Sugar > 120mg/dl

restecg Resting Electrocardiographic Results

oldpeak Previous Peak

exng Exercise Induced Angina

thall Thallium Stress Test Result

slp Slope



TARGET VARIABLE

"Output" Variable: Heart Attack Occurrence

- 1: Heart Attack Occurred
 - Indicates that the individual experienced a heart attack.
- 0: No Heart Attack
 - Indicates that the individual did not experience a heart attack.

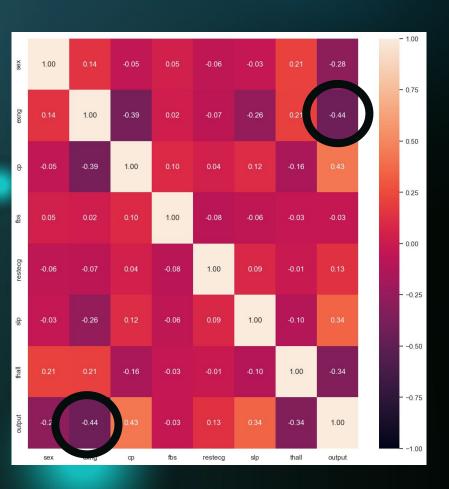
EXPLORATORY DATA ANALYSIS

- Data Cleaning
- Categorical visualization
- Numeric visualisation
- Feature selection

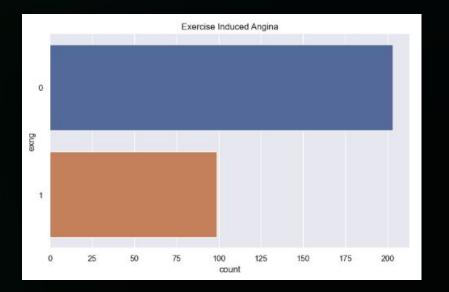


DATA CLEANING

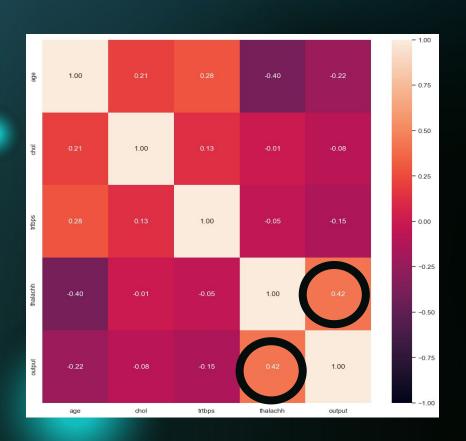
We found one duplicate value and hence have dropped it.

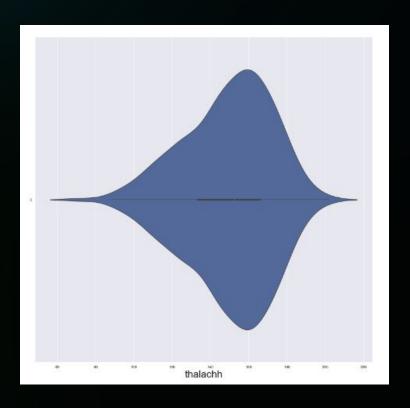


CATEGORICAL VISUALISATION



NUMERIC VISUALISATION



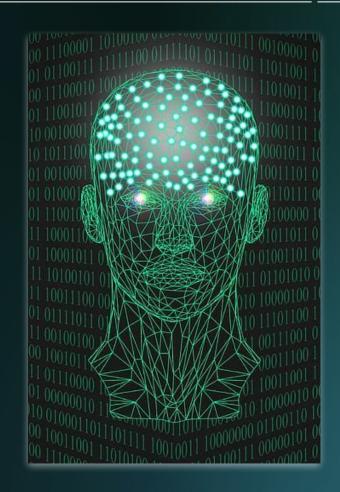


FEATURE SELECTION

```
Logistic Regression Accuracy: 0.8524590163934426
Logistic Regression Classification Report:
              precision
                           recall f1-score
                                               support
                   0.86
                             0.83
                                       0.84
           0
                                                    29
                   0.85
                             0.88
                                       0.86
                                                    32
                                       0.85
                                                    61
    accuracy
   macro avq
                   0.85
                             0.85
                                       0.85
                                                    61
weighted avg
                   0.85
                             0.85
                                       0.85
                                                    61
Feature Importance using Logistic Regression:
     Feature Coefficient
                -1.172798
         sex
12
       thall
              -1.142387
11
              -0.769140
         caa
8
              -0.730590
        exng
2
                 0.696413
          ср
10
                 0.656893
         slp
6
                 0.546694
     restecq
9
     oldpeak
                -0.480931
5
         fbs
                 0.076861
    thalachh
                 0.033457
3
      trtbps
                -0.021249
                 0.020281
         age
        chol
                -0.002199
```

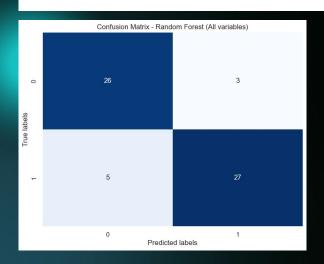
MACHINE LEARNING

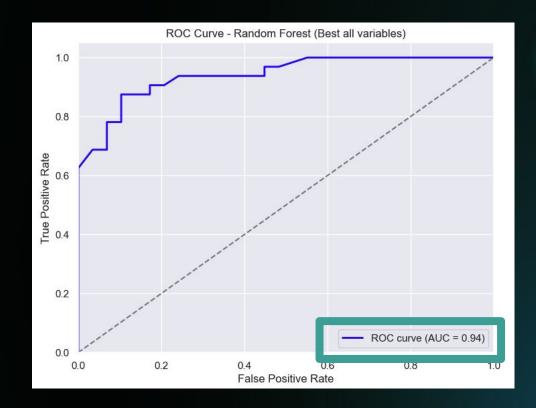
- Random Forest
- Gradient Boosting



RANDOM FOREST CLASSIFIER

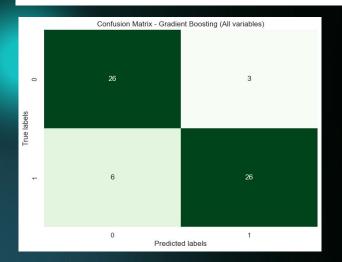
Random Forest Random Forest				88524590163934 iables):
	precision	recall	f1-score	support
0	0.84	0.90	0.87	29
1	0.90	0.84	0.87	32
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

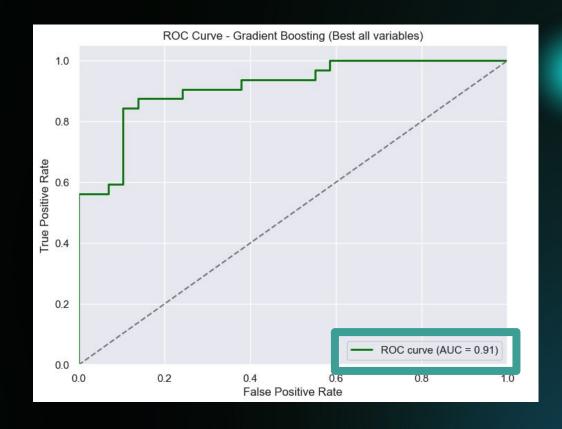




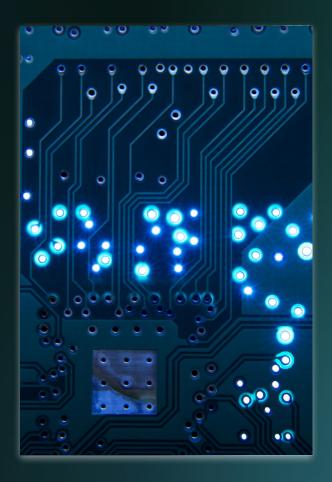
GRADIENT BOOSTING

Gradient Boosti Gradient Boosti p		cation Rep	ort (All v		6
0 1	0.81 0.90	0.90 0.81	0.85 0.85	29 32	
accuracy macro avg weighted avg	0.85 0.86	0.85 0.85	0.85 0.85 0.85	61 61 61	





FINAL INSIGHTS





FINAL OBSERVATIONS

- Both Random Forest and Gradient Boosting perform well.
- Models trained with fewer variables have slightly lower performance.
- Random Forest achieves higher AUC scores compared to Gradient Boosting.
- Precision and recall are well balanced across classes for most models.

COURSE OF ACTION

Through our exploratory data analysis, we have received multiple observations.

In order to come to a conclusion from our observation, we performed **correlation analysis**.

The analysis of all the variables allowed us to perceive the variable which has the highest correlation with the 'target' variable.

We note down those variables and **compare** with the feature selection.

Through the machine learning model of logistic regression, feature selection is performed. The second machine learning model, the Random Forest Classifier is implemented based on the feature selection The third machine learning model, the **Gradient Boosting** is implemented based on feature selection Comparison of the accuracy of the two models Therefore, coming to the conclusion which model is better.

THANKS!_

Do you have any questions?

CREDITS: This presentation template was created by <u>Slidesgo</u>, and includes icons by <u>Flaticon</u>, and infographics & images by <u>Freepik</u>

