

# Heart Stroke/Attack Analysis

-Vishnu Teja Mekala  
-Praveen Nakka  
-Vamshi Krishna  
Mandalapu

1

## Resources

- We intend on using python libraries like pandas,Scikit-Learn and Matplotlib while searching for the tutorials on Kaggle to better know the methods used by other data scientists in this field
- We plan on viewing previous research works in predicting heart disease using machine learning.

4

## Abstract

- The Goal of this project is to create a machine learning model which can predict the probability of a person getting a heart attack using several factors such as the patients age,sex,blood pressure.
- Heart diseases are one of the major health concerns which is leading to deaths worldwide,So this is to detect these attacks before they occur so that proper precautions can be taken.
- To identify the best model for this, we will explore several machine learning algorithms like logistic regression, SVMs and random forest. We will evaluate these models' performance using cross-validation techniques and ROC curves. By doing so, we aim to provide healthcare professionals with valuable insights into identifying individuals at risk of heart disease and taking preventive measures.

2

## Related Projects

- Kaggle Credit Card Approval Prediction Competition
- <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>
- Scikit-learn Machine Learning in Python
- <https://scikit-learn.org/stable>

5

## Data Description and Potential Statistical Tests

- The dataset has each of the patients medical history and their habits such as if the person smokes.
- We use the heart\_attack.csv dataset which has around 303 rows of patient data and 14 different attributes regarding the patients details as columns.
- We use statistical tests like regression models and hypothesis testing to search for the features that are related to heart diseases and the features which are to be used is determined to have the predictions.

3

## Data Specifications

The data in this contains 303 patients details as observations and has 14 features/attributes such as age,sex,chest pain type,blood pressure,cholesterol,blood sugar ,ECG results ,maximum heart rate achieved,exercise angina ST depression induced by exercise and thalassemia,segment number of major vessels.

6

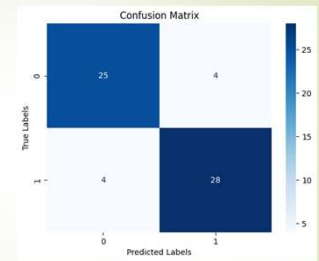
## List of features

- Age
- Sex
- Chest\_Pain\_Type
- Resting\_Blood\_Pressure
- Cholesterol
- Fasting\_Blood\_Sugar
- Resting\_ECG\_Results
- Max\_Heart\_Rate\_Achieved
- Exercise\_Induced\_Angina
- Old\_Peak
- Slope
- Num\_Major\_Vessels
- Thalassemia
- Heart\_Attack\_Prediction

7

## Support Vector Machine

- Confusion matrix using heat map



10

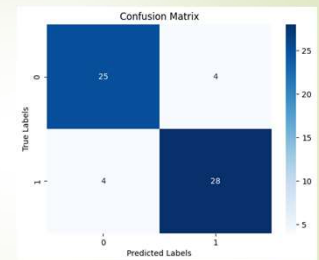
## Design and Milestones

- For the data preprocessing we fill in the missing values of the dataset by encoding categorical variables ,scaling numerical features before splitting them into the training and testing sets.
- We are building the model using several machine learning algorithms such as the Random forest ,SVM and Logistic Regression While recording their performances and determining their accuracy, precision recall and F1 score.
- We plan on having the accuracy to be 80% and we will also create visualizations and plots for interpreting the data.
- Data cleansing ,exploratory data analysis,tuning and final model selection, also deployment is included in our goals.

8

## Random Forest Matrix

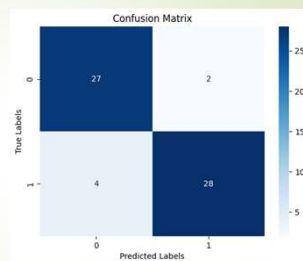
- Confusion matrix using heat map



11

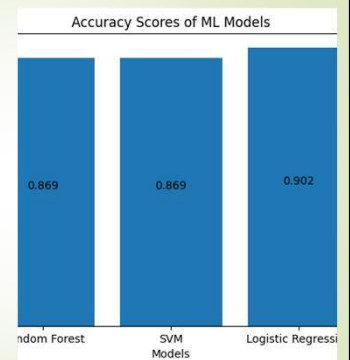
## Model Building

1. Logistic Regression:  
Confusion matrix plot using a heat map

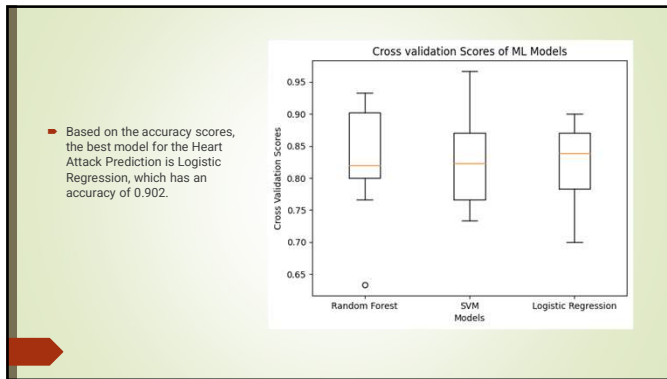


9

## Accuracy Scores of Machine Learning models



12



13

Repository/Archive

https://github.com/PraveenNakk/Emphirical-project1

14

Thank You

15