


ANALYSIS AND PREDICTION OF HEART ATTACKS



*By Group3:
Tala ALFuhaid
Ghadeer AlMansour
Najd AlNafisah
Jenan AlMeshari
Raghad Alzamil*



CONTENT

1

Our Problem

2

Dataset

3

**Machine
Learning
techniques**

4

**Results and
Findings**

PROBLEM



THE PROBLEM



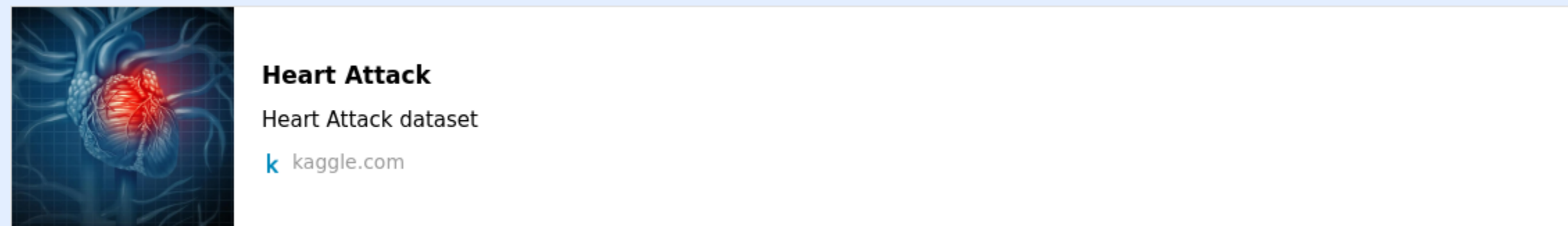
Heart attacks are a major health problem globally. Knowing who is at high risk for a heart attack isn't always easy because many different things, like age, lifestyle, and health, can affect it. Finding out early who might have a heart attack can save lives. We want to use machine learning to make a tool that can better predict who is at risk of a heart attack.

DATASET



OUR DATASET

The **Heart Attack** dataset is a collection of data related to the risk factors for heart attacks. The dataset was compiled by Kaggle and it includes demographic information such as age and gender, as well as medical information such as blood pressure and cholesterol level. The goal of the dataset is to predict the likelihood of a person having a heart attack based on various risk factors.



DATASET GENERAL INFORMATION

- 303 Observations
- 14 Variables, which include 13 patient attributes, and one target variable
- Types of variables in the dataset are:
 1. Binary
 2. Numeric
 3. Categorical.



DATASET GENERAL INFORMATION

BINARY PATIENT ATTRIBUTES:

- Sex
- exercise-induced angina
- fasting blood sugar > 120 mg/dl



DATASET GENERAL INFORMATION

NUMERIC PATIENT ATTRIBUTES:

- age
- resting blood pressure
- ST depression induced by exercise
- serum cholesterol level
- maximum heart rate achieved during exercise



DATASET GENERAL INFORMATION

CATEGORICAL PATIENT ATTRIBUTES:

- chest pain type
- slope of the peak exercise ST segment
- number of major vessels colored by fluoroscopy
- thallium stress test result
- electrocardiographic result



DATASET GENERAL INFORMATION

THE TARGET VARIABLE:

Our dataset contains a binary target variable indicating whether heart disease is present or absent. This variable can be termed as the class or label and offers two outcomes: label 0 for patients without heart disease and label 1 for those with it.



MACHINE LEARNING TECHNIQUES



MACHINE LEARNING TECHNIQUES

The primary objective of this project is to classify patients as either susceptible or not susceptible to heart attacks. To achieve this prediction, we will employ two distinct types of machine learning techniques:

- Supervised learning
- Unsupervised learning



MACHINE LEARNING TECHNIQUES

FOR SUPERVISED LEARNING

**LINEAR
REGRESSION**

DECISION TREE

MACHINE LEARNING TECHNIQUES

SUPERVISED LEARNING (LINEAR REGRESSION):

Linear regression, a regression technique employed in diagnosing heart attack susceptibility, will be used to make predictions about the patients if they are a prime target for heart attacks or not. This involves employing statistical procedures to estimate the relationship between the target variable (dependent variable) and independent variables such as: age, sex, Cp, Tresps, Chol, Fbs, restecg, thalach, exang, oldpeak, slope, ca, and thal.



MACHINE LEARNING TECHNIQUES

SUPERVISED LEARNING (DECISION TREE):

We will also implement decision trees, a classification methodology, in our project. The C4.5 algorithm will be employed to construct the decision tree, facilitating the classification of patients as either prime targets for heart attacks or not. This categorization is based on the how the previous questions were answered.



MACHINE LEARNING TECHNIQUES

UNSUPERVISED LEARNING

In the realm of unsupervised learning, we will employ K-Means, a clustering algorithm designed to identify similar observations within a dataset and group them into clusters. Each cluster will be associated with a centroid. The algorithm takes the number of clusters, denoted as K , and the dataset as inputs, and it produces the dataset partitioned into K clusters as outputs.



RESULTS AND FINDINGS



MATRICES:

LOGISTIC REGRESSION METRICS:

- ACCURACY = 75.41%
- ERROR RATE = 24.59%
- ROC AREA UNDER CURVE = 0.75
- PRECISION = 0.72
- RECALL = 0.84
- F1 = 0.78

DECISION TREE METRICS:

- ACCURACY = 77.05%
- ERROR RATE = 22.95%
- ROC AREA UNDER CURVE = 0.77
- PRECISION = 0.76
- RECALL = 0.81
- F1 = 0.78

PRECISION AND RECALL VALUES:

K=2:

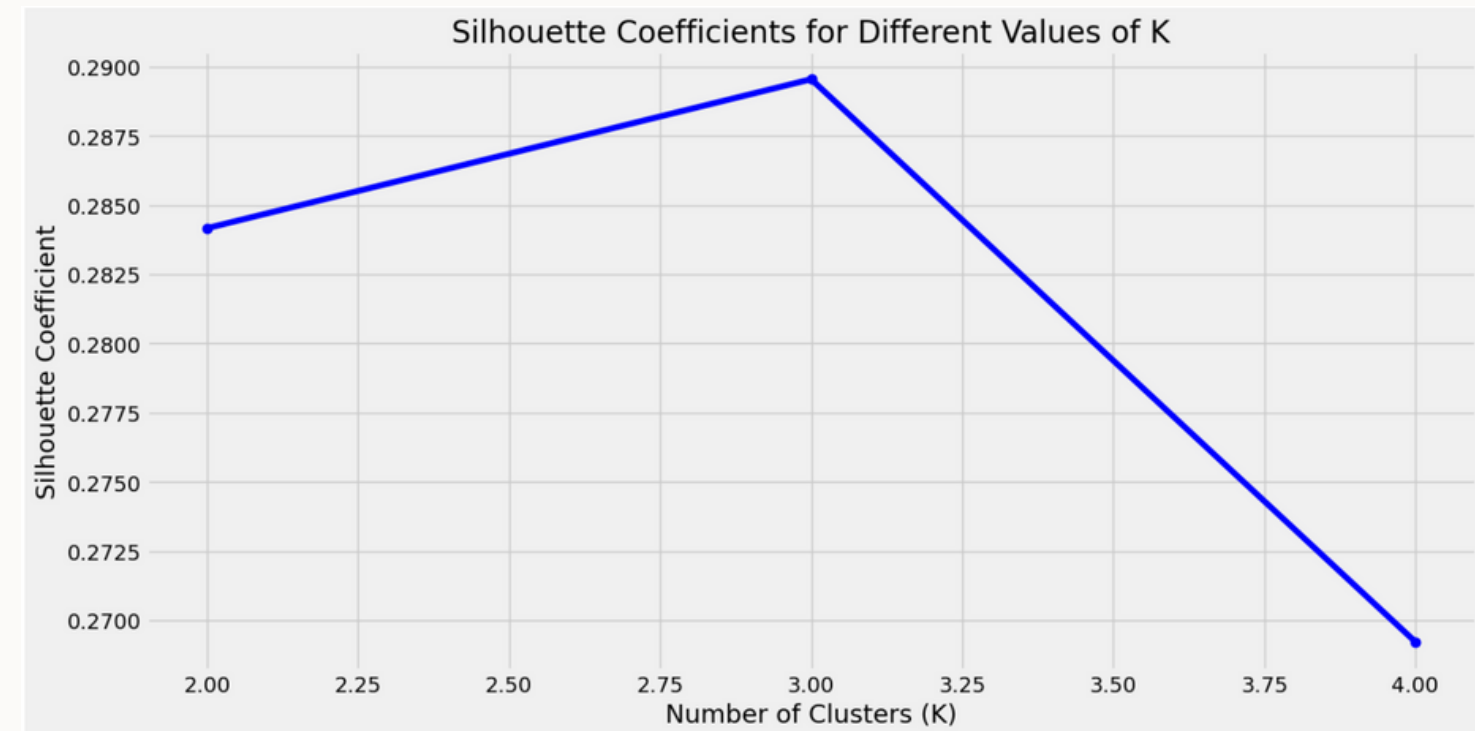
- PRECISION FOR K=2: 0.4388
- RECALL FOR K=2: 0.4290

K=3:

- PRECISION FOR K=3: 0.4391
- RECALL FOR K=3: 0.3630

K=4:

- PRECISION FOR K=4: 0.4071
- RECALL FOR K=4: 0.2211



**THANK
YOU!**



REFERENCES

[1] . [https://www.kaggle.com/datasets/pritsheta/heart-attack?
resource=download](https://www.kaggle.com/datasets/pritsheta/heart-attack?resource=download)

