

# CS 513 B – KDD PROJECT PROPOSAL: (Heart Attack Analysis)

CS 513 B - KDD Project Proposal: Heart Attack Analysis

## Project Group No: 5

### Problem Statement:

Heart Attack Analysis is a problem that involves predicting the likelihood of a person experiencing a heart attack based on various risk factors. The problem statement typically involves using a dataset of patient information, such as age, gender, cholesterol levels, blood pressure, and other health metrics, to develop a predictive model that can accurately identify individuals at high risk of a heart attack. The goal of this analysis is to enable healthcare professionals to identify patients who may require additional screening or treatment to prevent a potentially life-threatening heart attack.

### Dataset: (Description)

The dataset comprises 13 features in the form of columns, out of which we may opt to use the essential features only, during implementation.

Age : Age of the patient

Sex : Sex of the patient

exang: exercise induced angina (1 = yes; 0 = no)

ca: number of major vessels (0-3)

cp : Chest Pain type chest pain type

Value 1: typical angina

Value 2: atypical angina

Value 3: non-anginal pain

Value 4: asymptomatic

trtbps : resting blood pressure (in mm Hg)

chol : cholestoral in mg/dl fetched via BMI sensor

fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

rest\_ecg : resting electrocardiographic results

Value 0: normal

Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria  
thalach : maximum heart rate achieved

target : 0= less chance of heart attack 1= more chance of heart attack

Source of Dataset:

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

Implementation Strategy and algorithms used: (List different models)

We have decided to implement and compare 8 other models among four different group members. We have chosen a few models from our course and a few from outside the course. The following are the models selected by us:

1. K-nearest neighbor
2. AdaBoost Classifier
3. Random Forest
4. Support Vector Machine
5. Gaussian Naive Bayes
6. Logistic Regression
7. Decision tree
8. Artificial Neural network

Model metrics and Evaluation: (List evaluation metrics like AUC-ROC, Confusion matrix, F1, Recall, and Precision)

Evaluation of different models used in the project are

- Confusion Matrix
- F1
- Recall
- Precision

Team Members: Group 1

1. Member 1 Aman Gupta -20018346
2. Member 2 Parth Rajeshkumar Patel - 20010839
3. Member 3 Ashwin Dhanasamy
4. Member 4 Buruju Sowmya 20016102