

Heart Attack Analysis and Prediction

We will discuss the following points as they relate to my project:

1. Problem Identification
 2. Research and Justification
 3. Dataset Selection
 4. Methodology
 5. Team Recruitment
-

1. Problem Identification

Identify a specific issue or challenge that you are passionate about solving. The problem should be significant and relevant, offering a clear opportunity to apply machine learning/deep learning methods.

Public health is of central importance and one specific area of significance is cardiac health. Heart attacks are a major issue and the ability to accurately predict and prevent heart attacks is invaluable. Using large datasets, we can analyze the major contributing factors to heart attacks and build models to predict the likelihood of patients having a heart attack. This will allow us to take more preventative measures for high-risk patients.

So, the task at hand is to accurately predict the likelihood of a patient having a heart attack given relevant data about the patient.

2. Research and Justification

Conduct thorough research to understand the context and background of the problem. Provide a justification for why this issue is important and how machine learning can be leveraged to find a solution.

Heart attacks are an extremely common illness, far more common than they should be, affecting millions of people every year. Traditional methods to diagnosis patients for being at risk for a heart attack may not be as accurate as they could be. This opens the opportunity for machine learning algorithms to uncover patterns, that may not initially be obvious, by learning from large datasets. This can improve accuracy of predictions as well as enhance early detection of patient risk to heart disease. Early detection and improved accuracy can in turn lead to saving people's lives. The value of even a single life can never truly be understood. Therefore, this project is of immense importance.

3. Dataset Selection

Select an appropriate dataset that is well-suited to the problem you aim to solve. The dataset should be comprehensive, relevant, and of high quality. Ensure that the data is properly pre-processed and ready for analysis.

The “**Heart Attack Analysis & Prediction Dataset**” from Kaggle

(<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>) was selected. This is a high-quality dataset including many relevant features to help us accurately predict the likelihood of a patient’s risk of heart attack.

4. Methodology

Outline the machine learning techniques and algorithms you use. Provide a rationale for your choices and explain how they are appropriate for the problem at hand. Be sure to include validation and evaluation methods to measure the effectiveness of your solution.

Machine learning algorithms vary in terms of speed, complexity, accuracy and various other factors. Below is an outline of the machine learning algorithms that are used in the project:

- Logistic Regression
- Random Forest
- Gradient Boosting
- XGBoost
- Support Vector Machine
- Naïve Bayes
- K-Nearest Neighbors

The rationale behind the selection of these algorithms is that a Logistic Regression model can provide us with a simple baseline model to gauge baseline accuracy. It is appropriate for a binary classification problem, such as the one at hand, and gives us an estimate of the probability/likelihood of a heart attack. The additional algorithms will be used to further improve accuracy, and model more complex patterns in the data. Hyperparameter tuning was also done for the best performing models.

Prior to training the machine learning models, we apply exploratory data analysis (EDA) techniques to more deeply understand the data and trends. This includes both univariate as well as bivariate analysis.

Validation and evaluation are critical in a machine learning project. The data was split into two sets: (a) a training set and (b) a test set. Splitting the data this way allowed us to

evaluate our model properly. In terms of evaluation, five measures of effectiveness were used: (a) accuracy, (b) precision, (c) recall, (d) F1-score, and (e) ROC-AUC.

Lastly, two additional steps were taken to further enhance the project: (a) explainability and (b) interactive visualizations. We used SHAP (SHapley Additive exPlanations) values to indicate feature importance and to add explainability/interpretability to the best performing machine learning models. This allowed us to better understand the results of the models instead of leaving them opaque. Various interactive visualizations were also included using the Plotly library. This facilitated new analysis of the data that wasn't possible before, such as 3D scatter plots.

5. Team Recruitment

Write a compelling requirement to recruit more members for your project. Clearly describe the skills and expertise you are looking for in potential team members. Highlight the value and impact of the project to attract motivated and skilled individuals.

The goal of this project is to develop a machine learning model that can accurately predict the likelihood of a patient's risk of heart attack based on relevant patient data. The impact of this project will be worldwide, helping doctors around the globe better serve patients. The project will allow doctors to have improved early detection methods as well as more accurate predictions regarding a patient's health risks. Ultimately, this will lead to people's lives being saved, the value of which can never be understated.

The following skills and expertise are being sought out:

- Data processing: data cleaning, preprocessing, and experience with EDA (Exploratory Data Analysis). Experience with Pandas, Matplotlib, and Seaborn is also critical.
- Machine Learning Algorithms: experience with various machine learning algorithms is important, including: logistic regression, decision trees and other related algorithms. Python will be used along with libraries such as scikit-learn. Related skills include experience with hyper-parameter tuning and model evaluation.
- Communication skills: strong presentation and organization skills are necessary. Members must be able to work in a team environment.