HEART ATTACK PREDICTION

Madhavi Ghanta April 14, 2024
BELLEVUE UNIVERSITY
DSC 680 Spring 2024

TOPIC

It is critical to comprehend the dangers associated with heart attacks in the modern healthcare environment. Heart attacks continue to be a leading cause of death worldwide, and identifying those who may be more susceptible to them might help avert these potentially fatal situations. With its data-driven approach to heart attack risk prediction, this research hopes to shed light on cardiovascular health.

BUSINESS PROBLEM

Myocardial infarctions, another name for heart attacks, are a complex issue with broad ramifications for people's health as well as for the healthcare system as a whole. An overview of the main problems related to heart attacks is provided below:

- High Mortality Rate: Heart attacks rank among the world's leading causes of mortality. They can occur suddenly and fatally, and life frequently hinges on receiving treatment right away. Many people who suffer a heart attack either die or experience serious effects, even with advancements in therapy.
- Lifestyle-Related Risk Factors: Adopting unhealthy lifestyle habits, such as smoking, binge drinking, eating poorly, and not exercising, greatly raises the chance of heart attacks. These habits are part of the issue and are common in many societies.
- Aging Population: An increase in heart attack incidence is predicted as the world's population ages. Age is an unchangeable risk factor, and heart disease is more common among the elderly.
- Medical Expenses: Hospital stays, operations, prescription drugs, and long-term care are among the many medical expenses associated with heart attacks. The financial strain can be unbearable for individuals, insurers, and healthcare systems.

- Treatment Gaps: During a heart attack, having access to prompt and efficient care is essential. Inadequate healthcare facilities in rural or underdeveloped locations might cause treatment delays and subpar results.

A comprehensive strategy that emphasizes prevention, early detection, access to high-quality healthcare, and enhancing the general cardiovascular health of people and populations is needed to address this issue.

DATASET

This dataset is sourced from Kaggle(Fahad Mehfooz.). This dataset has around 14 features with around 300 rows. Following are the features available in this dataset.

- 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

Given the limited number of rows in the current dataset, should the model's outcomes exhibit inconsistency, I intend to explore an alternative dataset on Kaggle, one with a larger number of rows, approximately 900 (Fedesoriano).

METHODS

We will do a number of data preparation operations, such as addressing missing data, removing duplicates, etc., to guarantee the quality and applicability of the dataset before supplying it to the models. For this project, I intend to construct the following models using the cleaned dataset.

• **Logistic Regression:** Frequently employed for binary classification applications such as heart attack prediction (0 or 1), logistic regression is a straightforward and easily comprehensible model. It can shed light on the relative contributions of each feature to prediction.

• **Random Forest:** This resilient ensemble model effectively manages non-linearity and functions with both numerical and categorical features. It can produce feature importance rankings and extract complex relationships from the data.

• **Support Vector Machines (SVMs):** SVMs can handle both linear and non-linear data and are effective for binary classification. They come in particularly handy when working with data that has many dimensions.

• **Naive Bayes:** This probabilistic classifier can be computationally effective and performs well with categorical features. It's thought to be a wise decision when handling with discrete clinical data.

ETHICAL CONSIDERATIONS

Like any healthcare-related project, predictive modeling for heart attack prediction raises a number of ethical issues that must be properly addressed to guarantee the responsible and moral use of patient data and to put the needs of the patient first. Some important ethical considerations are as follows:

i. Ensuring that patient data used for model testing and training is anonymous and complies with applicable privacy requirements (such as HIPAA in the US) is crucial. Patients' informed consent must be sought and their data usage must be made clear to them before being used for research.

ii. Securing patient data from cyberattacks, unlawful access, and breaches requires the implementation of strong data security procedures.

iii. To make sure the project complies with legal and ethical requirements, collaboration with healthcare professionals and legal experts should be carried out.

CHALLENGES & ISSUES

Challenges:

• One major problem may be finding high-quality, properly labeled healthcare data. Forecasts that are not accurate or full can be deemed untrustworthy.

• It's possible that heart attack datasets are unbalanced, with a much higher proportion of patients (goal 0) than having had a heart attack (target 1). Models with bias may result from this.

• It takes a lot of resources and time to clinically validate the predictive model to guarantee its efficacy and safety in actual healthcare settings.

Issues:

Patients' health conditions vary over time, and for models to be current and accurate, they must adjust to changing data.

REFERENCES:

Fahad Mehfooz. *HeartAttack prediction with 91.8 % Accuracy, Kaggle* - https://www.kaggle.com/code/fahadmehfoooz/heartattack-prediction-with-91-8-accuracy/input?select=heart.csv

Fedesoriano - *Heart Failure Prediction Dataset, Kaggle* - https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data