Heart Attack Prediction

Aditya Sumbaraju

Bellevue University

DSC680: Applied Data Science

Dr. Brett Werner

February 13, 2022

https://github.com/adityasumbaraju/aditya_portfolio/tree/main/HeartAttackPrediction

Topic

This project aims to predict Cardiovascular disease (a heart failure) from the clinical parameters of a given patient. I would be considering the Heart disease data set from UCI Machine learning Repository.

Business Problem

Heart attack is the number 1 cause of death compared to other diseases globally, taking an approximate estimation of 18 million lives each year, accounting for 31% of worldwide deaths. Heart failure can be prevented by addressing behavioral risk factors such as unhealthy diet, tobacco use, obesity (overweight concerns), physical inactivity, and heavy use of alcohol using population-wide strategies. If these risk factors are coupled with early treatment, it dramatically impacts its prognosis.

It is undoubtedly a difficult diagnosis to identify high-risk patients because of several multi-factorial contributory risk factors such as high blood pressure, diabetes, and high cholesterol. Here comes the need for machine learning and data mining to study, evaluate and predict the disease beforehand.

Medical researchers, doctors, and scientists are still contributing to machine learning (ML) techniques to develop interactive GUIs to predict the early detection of this disease. This is because of their superiority in classification compared to other traditional statistical approaches and pattern recognition. In this use case, I will be addressing below research questions.

Research Questions:

- Can physicians will be able to predict Cardiovascular disease with the help of patient demographics
- Does this prediction reduce the risk and prevent heart attack disease. Is early detection of heart attack possible?

Datasets

The dataset was gathered from the Machine Learning Repository from the Center for Machine Learning and Intelligent Systems at the University of California, Irvine. This directory contains four datasets concerning heart failure diagnosis. Features are numeric-valued. The data was collected from below mentioned four locations:

- University Hospital, Zurich, Switzerland (Switzerland.data)
- Cleveland Clinic Foundation (Cleveland.data)
- Hungarian Institute of Cardiology, Budapest (Hungarian.data)
- V.A. Medical Center, Long Beach, CA (long-beach-va.data)

All four database has the same format. The databases have 76 raw attributes; only 14 of them are used.

Metadata:

1. age – Age in Years
2. sex – sex(1-male;0=female)
3. cp- (chest pain type)
    - Value 1: typical angina
    - Value 2: atypical angina
    - Value 3: non-anginal pain
    - Value 4: asymptomatic
4. trestbps - resting blood pressure (measured in **mm Hg** on admission to the hospital)
5. chol - serum cholestoral in mg/dl
6. fbs - fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg - resting electrocardiographic results
    - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
    - Value 1: having ST-T wave abnormality (T wave inversions and ST elevation or depression of > 0.05 mV)
    - Value 0: normal
8. exang - exercise-induced angina (1 = yes; 0 = no)
9. thalach - maximum heart rate achieved
10. slope - the slope of the peak exercise ST segment
    - Value 1: upsloping
    - Value 2: flat
    - Value 3: downsloping
11. ca - number of major vessels (0-3) colored by fluoroscopy

12. oldpeak - ST depression induced by exercise relative to rest
13. thal - Thalium Heart Scan - thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14. num - the predicted attribute

Dataset details:

```
In [13]: hap_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 929 entries, 0 to 928
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       929 non-null    float64
 1   sex       929 non-null    float64
 2   cp        929 non-null    float64
 3   trestbps  871 non-null    float64
 4   chol      922 non-null    float64
 5   fbs       847 non-null    float64
 6   restecg   928 non-null    float64
 7   thalach   875 non-null    float64
 8   exang     875 non-null    float64
 9   oldpeak   867 non-null    float64
 10  slope     810 non-null    float64
 11  ca        605 non-null    float64
 12  thal      707 non-null    float64
 13  num       929 non-null    int64
dtypes: float64(13), int64(1)
memory usage: 101.7 KB
```

Sample Data:

```
In [17]: hap_df.head()
```

Out[17]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
0	53.0	1.0	4.0	123.0	282.0	0.0	0.0	95.0	1.0	2.0	2.0	2.0	7.0	1
1	52.0	1.0	4.0	165.0	0.0	NaN	0.0	122.0	1.0	1.0	1.0	NaN	7.0	1
2	60.0	1.0	4.0	132.0	218.0	0.0	1.0	140.0	1.0	1.5	3.0	NaN	NaN	1
3	51.0	1.0	4.0	140.0	0.0	0.0	0.0	60.0	0.0	0.0	2.0	NaN	3.0	1
4	63.0	1.0	4.0	140.0	187.0	0.0	2.0	144.0	1.0	4.0	1.0	2.0	7.0	1

Dataset source: https://archive.ics.uci.edu/ml/datasets/heart+disease

Methods

I would use CRISP-DM to build a heart attack prediction model. Below are the phases I would be targeting to achieve a better model in this methodology.

- Data collection
- Data preparation and preprocessing
- Modeling and testing
- Model deployment and monitoring

Ethical Considerations

Appropriate informed consent is fundamental to the ethical conduct of research in humans. Society has demanded more outstanding efforts to protect the individual rights of patients and human subjects. This is an evolving and complex area. To deal with the current

regulatory environment, we must understand and appreciate the historical basis for society's concerns, including physician authority's factual and perceived nature.

Medical Researchers and organizations are well-advised to carefully consider the basis for increasing ethical considerations in conducting research in humans and become familiar with regulations that must be met. Analysts, Scientists, and personnel interacting with patients and volunteer subjects should also understand acquiring consent and credentialing to document their understanding of the issues in obtaining consent from patients, dealing with conflicts of interest, and managing PII data.

Challenges/Issues

1. Anticipating whitespaces in data and need to work on data alignment.
2. Incorrect variable types.
3. Python package-related issues.
4. Inaccurate or messy patient demographics

Acknowledgment:

I want to thank the authors mentioned below for providing the dataset.

- Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
- University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
- University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
- V.A. Medical Center, Long Beach, and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

References

Frye, R. L., Simari, R. D., Gersh, B. J., Burnett, J. C., Brumm, S., Myerle, K., Jaffe, A. S., Holmes, D. R., Lerman, A., & Terzic, A. (2009, November 24). *Ethical issues in cardiovascular research involving humans*. Circulation. Retrieved February 10, 2022, from https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.107.752766#d3e854

Shah, D., Patel, S., & Bharti, S. K. (2020, October 16). *Heart disease prediction using machine learning techniques - S.N. computer science*. SpringerLink. Retrieved February 10, 2022, from https://link.springer.com/article/10.1007/s42979-020-00365-y

Foster, N. (2018, July 23). *The future of heart attack prediction*. Mended Hearts. Retrieved February 10, 2022, from https://mendedhearts.org/story/the-future-of-heart-attack-prediction/

C. Thirumalai, A. Duba and R. Reddy, "Decision making system using machine learning and Pearson for heart attack," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 206-210, doi: 10.1109/ICECA.2017.8212797.

S. Manikandan, "Heart attack prediction system," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, pp. 817-820, doi: 10.1109/ICECDS.2017.8389552.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.