Visualizing and Predicting Heart Diseases with an Interactive Dash Board

Team ID:PNT2022TMID14156

Faculty mentor: Raghuvaran S

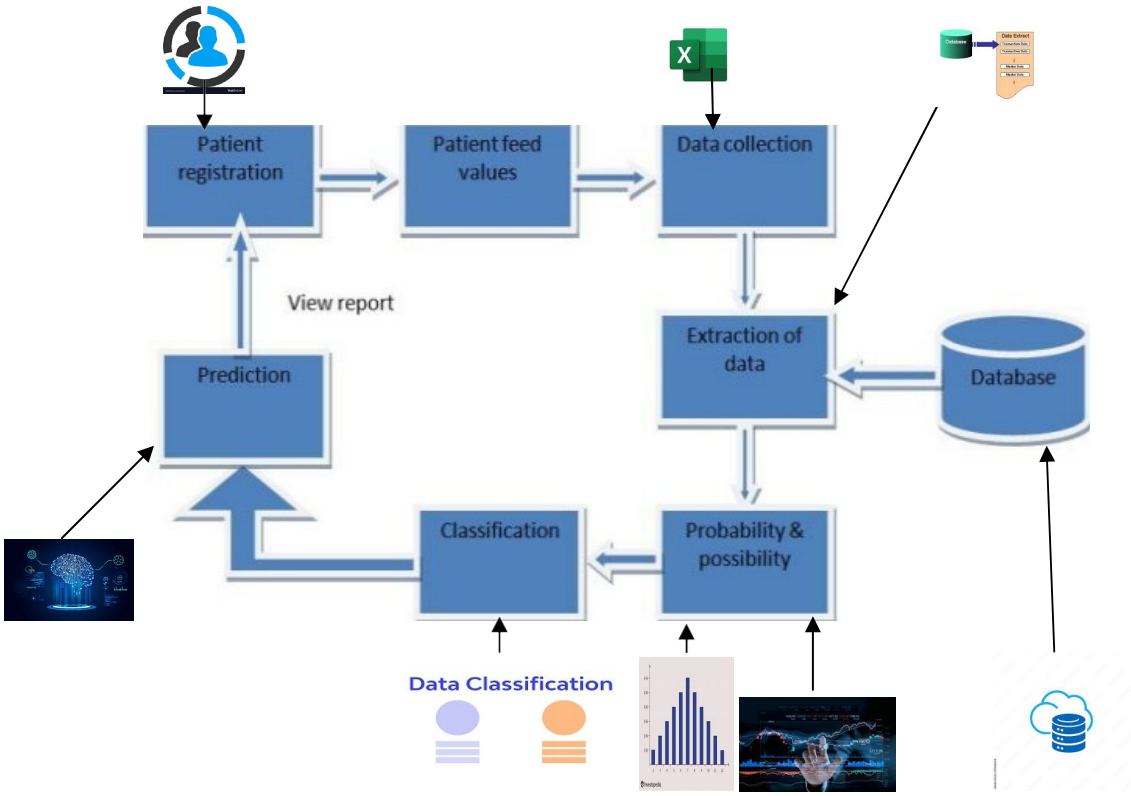
Team Leader: Vignesh

Team member : Yazhini

Team member: Suraksha

Team member : Shamyukta

SOLUTION ARCHITECTURE:



Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you're born with (congenital heart defects), among others.

The term "heart disease" is often used interchangeably with the term "cardiovascular disease". Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscles, valves or rhythm, also are considered forms of heart disease.

Heart disease is one of the biggest causes of morbidity and mortality among the population of world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions.

The dataset consists of 303 individuals data. There are 14 columns in the dataset, which are described below.

1. Age: displays the age of the individual.
2. Sex: displays the gender of the individualusing the followingformat :
   1 = male
   0 = female
3. Chest-pain type: displays the type of chest pain experienced by the individual using the following format:
   1 = typical angina
   2 = atypical angina
   3 = non-anginal pain
   4 = asymptotic
4. Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholesterol: displays the serum cholesterol in mg/dl (unit)
6. Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then:
   1(true) else : 0 (false)
7. Resting ECG: displays resting electrocardiographic results
   0 = normal
   1 = having ST-T wave abnormality
   2 = left ventricular hypertrophy
8. Max heart rate achieved: displays the max heart rate achieved by an individual.
9. Exercise-induced angina :
   1 = yes
   0 = no
10. ST depression induced by exercise relative to rest: displays the value which is an integer or floats.
11. Peak exercise ST segment :
   1 = upsloping
   2 = flat
   3 = downsloping
12. A number of major vessels (0–3) coloured by fluoroscopy : displays the value as integer or float.

13. Thal : displays the thalassemia :
    3 = normal
    6 = fixed defect
    7 = reversible defect
14. Diagnosis of heart disease: Displays whether the individual is suffering from heart disease or not:
    0 = absence
    1, 2, 3, 4 = present.

In the actual dataset, we had 76 features but for our study, we chose only the above 14 because :

1. Age: Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. It is estimated that 82 per cent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.
2. Sex: Men are at greater risk of heart disease than pre-menopausal women. Once past menopause, it has been argued that a woman's risk is similar to a man's although more recent data from the WHO and UN disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.
3. Angina (Chest Pain): Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood. It may feel like pressure or squeezing in your chest. The discomfort also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.
4. Resting Blood Pressure: Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.
5. Serum Cholesterol: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of a heart attack.
6. Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of a heart attack.

7. Resting ECG: For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercising ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.
8. Max heart rate achieved: The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.
9. Exercise-induced angina: The pain or discomfort associated with angina usually feels tight, gripping or squeezing, and can vary from mild to severe. Angina is usually felt in the centre of your chest but may spread to either or both of your shoulders, or your back, neck, jaw or arm. It can even be felt in your hands. o Types of Angina a. Stable Angina / Angina Pectoris b. Unstable Angina c. Variant (Prinzmetal) Angina d. Microvascular Angina.
10. Peak exercise ST segment: A treadmill ECG stress test is considered abnormal when there is a horizontal or down-sloping ST-segment depression ≥ 1 mm at 60–80 ms after the J point. Exercise ECGs with up-sloping ST-segment depressions are typically reported as an ‘equivocal’ test. In general, the occurrence of horizontal or down-sloping ST-segment depression at a lower workload (calculated in METs) or heart rate indicates a worse prognosis and a higher likelihood of multi-vessel disease. The duration of ST-segment depression is also important, as prolonged recovery after peak stress is consistent with a positive treadmill ECG stress test. Another finding that is highly indicative of significant CAD is the occurrence of ST-segment elevation > 1 mm (often suggesting transmural ischemia); these patients are frequently referred urgently for coronary angiography.

The Approach

The code for this article can be found here. The code is implemented in Python and different classification models are applied.

In this article I will be using the following classification models for classification :

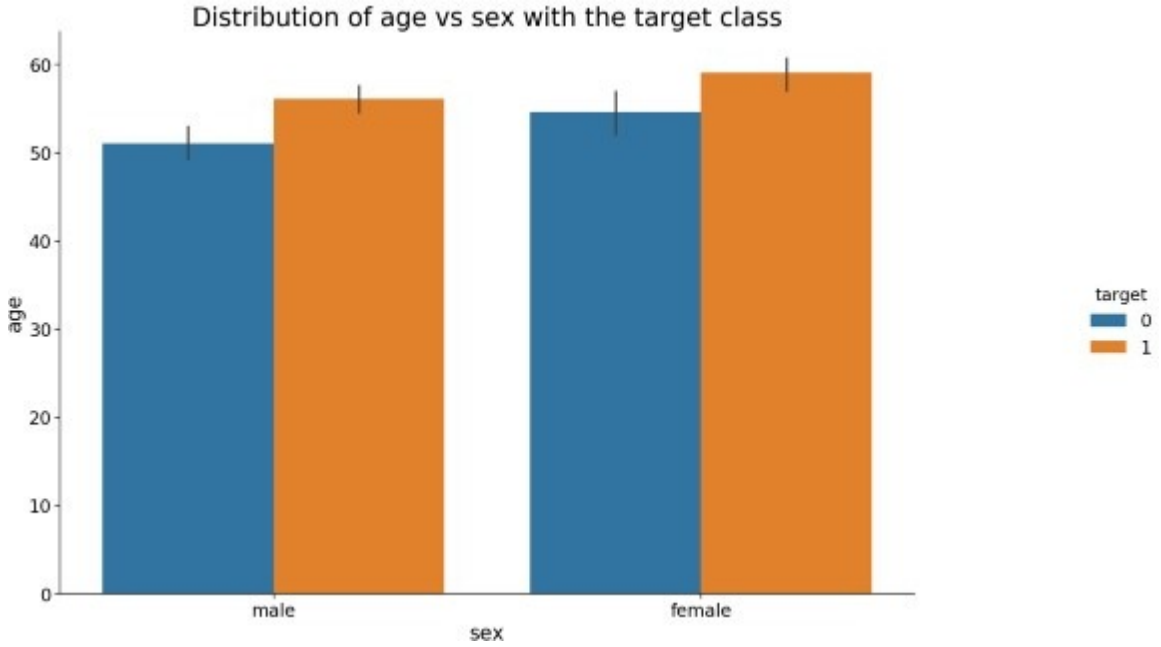
- SVM
- Naive Bayes
- Logistic Regression
- Decision Tree
- Random Forest
- LightGBM
- XGboost

Data Analysis

Let us look at the people's age who are suffering from the disease or not. Here, target = 1 implies that the person is suffering from heart disease and target = 0 implies the person is not suffering.

We see that most people who are suffering are of the age of 58, followed by 57. Majorly, people belonging to the age group 50+ are suffering from the disease.

Next, let us look at the distribution of age and gender for each target class.



We see that females who are suffering from the disease are older than males.

The Data

The dataset used in this article is the Cleveland Heart Disease dataset taken from the UCI repository.

Index	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
5	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
6	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
7	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
8	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
9	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
10	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
11	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
12	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
13	44	1	2	120	263	0	0	173	0	0	1	0	7	0
14	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
15	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
16	48	1	2	110	229	0	0	168	0	1	3	0	7	1
17	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
18	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0

Advantages

- Users can search for a doctor's help at any point in time.
- Users can talk about their Heart Disease and get an instant diagnosis.
- Doctors get more clients online.
- Very useful in case of emergency.

Disadvantages

- Accuracy Issues: A computerized system alone does not ensure accuracy, and the warehouse data is only as good as the data entry that created it.
- The system is not fully automated, it needs data from the user for a full diagnosis

The Heart Disease prediction will have the following key takeaways:

1. **Data insight:** As mentioned here we will be working with the heart diseasedetection dataset and we will be putting out interesting inferences from the data to derive some meaningful results.
2. **EDA:** Exploratory data analysis is the key step for getting meaningfulresults.
3. **Feature engineering:** After getting the insights from the data we have toalter the features so that they can move forward for the model-building phase.
4. **Model building:** In this phase, we will be building our Machine learningmodel for heart disease detection.

Feature information of the Cleveland dataset.

S.No	Attribute Name	Description	Range of Values
1	Age	Age of the person in years	29 to 79
2	Sex	Gender of the person [1: Male, 0:Female]	0, 1
3	Cp	Chest pain type [1-Typical Type 1Angina 2- Atypical Type Angina3-Non-angina pain 4-Asymptomatic)	1, 2, 3, 4
4	Trestbps	Resting Blood Pressure in mm Hg	94 to 200
5	Chol	Serum cholesterol in mg/dl	126 to 564
6	Fbs	Fasting Blood Sugar in mg/dl	0, 1

7	Restecg	Resting Electrocardiographic Results	0, 1, 2
8	Thalach	Maximum Heart Rate Achieved	71 to 202
9	Exang	Exercise Induced Angina	0, 1
10	OldPeak	ST depression induced by exerciserelative to rest	1 to 3
11	Slope	Slope of the Peak Exercise ST segment	1, 2, 3
12	Ca	Number of major vessels colored byfluoroscopy	0 to 3
13	Thal	3 – Normal, 6 – Fixed Defect, 7 –Reversible Defect	3, 6, 7
14	Num	Class Attribute	0 or 1

Patients from ages 29 to 79 have been selected in this dataset. Male patients are denoted by a gender value of 1 and female patients are denoted by a gender value of 0. Four types of chest pain can be considered indicative of heart disease. Type 1angina is caused by reduced blood flow to the heart muscles because of narrowed coronary arteries. Type 1 Angina is chest pain that occurs during mental or emotional stress. Non-angina chest pain may be caused due to variousreasons and may not often be due to actual heart disease. The fourth type,
Asymptomatic, may not be a symptom of heart disease. The next attribute *treetops* are the reading of the resting blood pressure. *Chol* is the
cholesterol level. *Fbs* is the fasting blood sugar level; the value is assigned as 1 if the fasting blood sugar is below 120 mg/dl and 0 if it is above. *Restecg* is the resting electrocardiographic result, *thalach* is the maximum heart rate, *exang* is exercise-induced angina which is recorded as 1 if there is pain and 0 if there is no pain, *oldpeak* is the ST depression induced by exercise, the *slope* is the slope of the peak exercise ST segment, *ca* is the number of major vessels colored by fluoroscopy, *thal* is the duration of the exercise test in minutes, and *num* is theclass attribute. The class attribute has a value of 0 for normal and 1 for patients diagnosed with heart disease.