Heart Failure Prediction using Machine Learning

DAB304-Healthcare Analytics Fall 2022 - 002
St. Clair College of Applied Arts and Technology, Windsor, ON N9G 3C3

Project Proposal

Introduction

Every year, approximately 17 million people are killed by cardiovascular diseases, which primarily demonstrate as cardiac and heart failures. It is a condition in which the heart is unable to pump enough blood to meet the body's requirements. Heart disease symptoms can include physical weakness, breathing problems, swelling feet, etc. The techniques are critical for identifying complicated heart diseases that pose a high risk to human life.

Healthcare industries generate massive amounts of data, known as big data, which contains hidden knowledge or patterns for decision making. Data analysis is crucial in healthcare, including new research findings, emergency situations, and epidemics. Analytics in healthcare improves care by enabling preventive care, and Exploratory Data Analysis (EDA) is a key step in data analysis. Moreover, A Machine Learning (ML) model can be very helpful in the early detection and care of people with cardiovascular disease or at high cardiovascular risk (due to the presence of one or more risk factors including like hypertension, diabetes, or pre-existing disease).

To predict heart disease, we will use multiple Machine learning models (Random Forest, Decision Tree, etc.) in conjunction with data analytics and visualisation tools. Therefore, there is a perfect fit between this project and this course.

Motivation

The delivery of superior services at low cost is a major challenge for healthcare groups (hospitals and medical centres). Quality service entails correctly diagnosing patients and administering effective treatments. Clinical tests must be kept as inexpensive as possible in hospitals. They can accomplish these goals by utilizing proper user information and/or decision support systems.

Today, most hospitals use hospital information systems to manage their healthcare or patient data. Patients' electronic medical records quantify symptoms, body features, and clinical laboratory test values, which can be used to perform biostatistics analysis to highlight patterns and correlations that medical doctors would otherwise overlook. Alas, these data are rarely used to help clinicians make clinical decisions.

These data contain a wealth of hidden information that is largely untapped. This begs the question, "How can we transform data into insights that will allow healthcare practitioners to make clinical decisions?"

Evaluation

The dataset holds 299 records with 13 attributes such as,

age
anaemia - *Decrease of red blood cells or hemoglobin (boolean)*
creatinine_phosphokinase - *Level of the CPK enzyme in the blood (mcg/L)*
diabetes - *If the patient has diabetes (boolean)*
ejection_fraction - *Percentage of blood leaving the heart at each contraction (percentage)*
high_blood_pressure - *If the patient has hypertension (boolean)*
platelets - *Platelets in the blood (kiloplatelets/mL)*
serum_creatinine - *Level of serum creatinine in the blood (mg/dL)*
serum_sodium - *Level of serum sodium in the blood (mEq/L)*
sex - *Woman or man (binary)*

smoking - *If the patient smokes or not (boolean)*
time - *Follow-up period (days)*
death_event - *If the patient deceased during the follow-up period (boolean)*

The project will be noted "successful" if it meets the following objectives:

— An ML Model will be deployed to a webpage and accessible to the public to predict the probability of heart failure by entering the above attributes.
— Achieving a prediction accuracy of between 80 to 90 for each of the models employed in the project to win the public's trust and confidence.
— A site with an integrated interactive analytics dashboard to promote awareness of trends and coronary heart disease risk factors.

Resources

Heart failure clinical records Data
https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records

Computational Tools : **Deep Notes, Tableau, Excel, PyCharm IDE, PowerBI**

Contribution

Below is a breakdown of how the our teamwork will be distributed according to project phases. The team intends to work discretely throughout the whole project.

	Data Collection	**Preprocessing & Cleaning**	**EDAs**	**Visualization & Dashboards**	**Apply ML Models**	**Deployment**
Smit Hareshkumar Rana **[792056]**	✓	✓		✓	✓	✓
Harsh Girishkumar Patel **[791820]**	✓		✓		✓	
Rechel Thomas Rebello **[787548]**	✓		✓	✓		
Dhruvkumar Limbachiya [**784687]**	✓	✓		✓		

References

[1] https://towardsdatascience.com/3-ways-to-deploy-machine-learning-models-in-production-cdba15b00e

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4216425/