**PREDICTING HEART ATTACK USING MACHINE LEARNING ALGORITHMS**

Department of Data Science  
Name: Sirajo Shehu  
ID: 14204294  
Email address: shehus@coventry.ac.uk

*Abstract*: now a days sudden death as result of heat attack has become common for both developed and developing countries studies of human inheritance can provide useful information for the development of advance technics by identifying people that may be at higher risk for heart attack. This research paper tends to show a precise result that accurately predicting the likelihood of heart attack based on previous risk factors and medical history, by using machine learning algorithms, this will be done by identify some pattern and tendency in data that may not be apparent to human predictor this has allowed for more personalised and proactive measures to prevent heart attack through early detection.

**Introduction:**

Heart is among the most important parts of human’s organs it requires very important care especially for people above sixty years. According to World Health Organization an estimated 17.9 million lives died because of cardiovascular diseases and more than four out of CVD deaths are due to heart attacks and stroke which happens prematurely (WHO, 2020). Researchers engage exploring information through the application of machine learning classifiers[]. Numerous academic articles have inquired into predicting the presence of heart disease by both academics and medical practitioner. Various methodologies and studies have emerged over time to categorize heart disease using data mining and machine learning techniques[]. The authors aim to provide a thorough review of research regarding the application of machine learning in the context of heart conditions. They propose a dataset containing the basic samples and data needed for developing an effective method for predicting heart diseases. Efficient preprocessing of the dataset is imperative to ensure its suitability for use by machine learning algorithms, eventually leading to advanced predictive conclusions.

In recent time various approach of machine learning algorithms were used to predict the tendencies of heart attack. Some of the research conducted includes predicting emergency readmission of heart attack patients that requires urgent attention [2]. Also, a thorough review of machine learning method of Stacking Ensemble Learner (SEL) was developed [2]

**Problem and Data Set Description:**

The challenge many cardiovascular disease (CAD) patients are facing is that inadequate knowledge of been prone to heart attack depending on the degree of the contributing factors, this research tends to identify these.

This data set was obtained from Kaggle website, and it has 14 variables (columns) with patients record of 303 the variables are outline below:

1. Age
2. Sex, exang which is exercise induced angina(1=yes; 0=no);
3. Coronary Artery (ca)
4. Chest pain( cp):
5. Resting blood pressure (trtbps):
6. Cholesterol (chol): cholesterol in mg/dl fetched via BMI sensor;
7. Fasting Blood Sugar fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false);
8. Resting Electrocardiographic (restecg): 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);
9. Maximum Heart Rate Achieved (thalach): maximum heart rate achieved; target: 0=less chance of heart attack 1= more chance of heart attack.
10. Exercise-Induced Angina (exng)
11. oldpeak
12. Slope
13. Thalassemia (thal)
14. Target (output)

this research model uses K-Nearest Neighbours (K-NN), Decision Tree (DT) and Naive Bayes to predict which can be utilised for detection of having the chances of heart attack.

**Methods: Proposed algorithms:**

Support Vector SVC: Is a supervised learning algorithm that improves the versatility of machine learning by minimizing structured risk and has been widely used in the classification field. The Support vector classification (SVC) is a two-class model whose learning strategy is to maximize the interval and eventually transform it into a solution of convex quadratic programming. Assuming the total number of datasets is N, arrange the collected data as input-output pairs, which can be expressed as (xi,yi), ( , i = 1,2, …,N. xi is an input vector that includes factors such as the environment and the individual; is an output variable that includes only the individual's thermal sensation in the environment. Suppose yi = 1, a positive class, which is the first type of thermal sensation; yi = −1, a negative class, which is another type of thermal sensation.[5]

**Decision Tree:** Decision tree learning is a supervised learning approach used in statistics, data mining and machine learning. In this formalism, a classification or regression decision tree is used as a predictive model to draw conclusions about a set of observations. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. More generally, the concept of regression tree can be extended to any kind of object equipped with pairwise dissimilarities such as categorical sequences.

The Naive Bayes classifier utilises the tasks of classification to apply the concepts of the Bayesian theorem. It functions on the assumption that the existence of one characteristic does not imply the existence of any other feature. Because of its efficacy, this approach is frequently employed in classification problems. Bayes' Theorem is used by the Naive Bayes approach to figure out how likely a hypothesis is based on what we already know about conditions that might be linked to that hypothesis. It performs well even in non-linearly separable scenarios and is remarkably efficient for linearly separable problems. The mathematical expression for Bayes' Theorem is given below:

**Information:**



P (A|B) = P(B|A)P(A)/P(B)

A : Data with unknown classes

B : Hypothesis data A is a specific class

P (B|A) : The probability of the hypothesis is based

P (H) : Hypothesis probability A

P(B|A) Probability B based on conditions on hypothesis A

P(B) : Probability

Limitation and gaps: there is need for comprehensive and more diverse dataset to improve the algorithms. Additionally, there was lack of standardization in the approach used to the develop these algorithms making it difficult to compare result throughout studies. Future investigates or research should focus on addressing issues associated to data privacy and ethical concern as well as ensuring that the algorithms are transparent and explainable for health workers and patients.

Experimental setup:

The extracted data set submitted was clean up and it found that there were no missing values but its observed that there are data irregularities. It is not possible for a living being to have: 0 RestingBP, 0 Cholesterol, 0 FastingBS therefore, Exploratory Data Analysis was performed to understand the data in a better sense, different attributes were used to understand the correlation machine learning classifier was used to evaluate the pre-processed data set the classifiers display different precisions expected at the predicting of the heart disease. From the analysis it is observed that Patient with flat and down-slopping are more likely to have heart disease than patient with up-slopping.

Furthermore, patients

Patients with Angina, old Peak of 1.0 and above, have higher chances, and Patients with Fasting Blood Sugar are more likely to have heart disease,

StandardScaler was used to preprocessing data and after the above stage, the data was split into train and test split, 80 percent was use as train data while the remaining 20 percent as test data.

The data was then labelled as

EXPLORATORY DATA ANALYSIS

Results

Patient with flat and down-slopping are more likely to have heart disease than patient with up-slopping

Discussion and conclusions

K-NN produces the best highest result both on accuracy and predicting scores among the three algorithms, and we did not lose any information and its training is very fast:

**Social considerations**

The deployment of machine learning algorithms for predicting the chances of heart attacks can significantly impact public health. These solutions possess the capacity to save life and ease the responsibility on medical personnel by supporting prompt detection and assistance.

Notably to acknowledge the possible risk of worsening health biases if these technologies are not completely available across all demographic groups. A bias against black individuals was observed in an algorithm used inside healthcare organizations. This highlights the need of employing diverse dataset for training machine learning models as a means of reducing the recurrence of present health imbalances [6].

Ethical considerations

using machine learning causes concern about consent, accountability, and privacy. The accuracy of ML prediction is very important as false positives can generate redundant anxiety and medical interventions, while false positives can result in wasted chances for early treatment. There is also need for thorough validation and testing of ML model before clinical deployment. Patient must also be told about how their data is used and the role of Artificial Intelligence in their care, positioning with the principle of informed consent. Data security and patient confidentiality is important to maintain because models require large amount of personal health data in accordance with General Data Protection Data Protection Regulation (GDPR) which includes right of details of the algorithm’s prediction.

Legal considerations

There is need for legal frameworks between data scientist and health workers because using machine learning involves complicated regulation reform which is clearly stated by GDPR’s principle and liabilities should be considered in case of algorithms failure.

Professional considerations

The use of machine learning algorithms in the diagnosis of heart attacks requires medical professionals to maintain a prime balance between applying technological progress and upholding their medical knowledge. It is important to provide professionals with training to comprehend the capabilities and constraint of these tools supporting them to incorporate them into clinical procedure without unnecessary dependence, this requires the acquisition of both technical expertise and ethical knowledge in order to successfully navigate the moral workings brought about by machine learning technologies.

References:

[1] WHO, No Title, Top 10 Causes Death, 2020

[2] A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients Alireza Ghasemieh, Alston Lloyed, Parsa Bahrami, Pooyan Vajar, Rasha Kashef

[3] A. Rajkumar, G.S. Reena, Diagnosis of heart disease using datamining algorithm,

Glob. J. Comput. Sci. Technol. 10 (10) (2010) 38–43.

[4] Study of thermal sensation prediction model based on support vector classification (SVC) algorithm with data preprocessing.

[5] Direct Uncertainty Prediction for Medical Second Opinions

Maithra Raghu, Katy Blumer, Rory Sayres, Ziad Obermeyer, Bobby Kleinberg, Sendhil Mullainathan, Jon Kleinberg Proceedings of the 36th International Conference on Machine Learning, PMLR 97:5281-5290, 2019.