**PREDICTING HEART ATTACK USING MACHINE LEARNING ALGORITHMS**

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

*Abstract*: (M. Rizwan, S. Arshad, H. Aijaz, R. A. Khan and M. Z. U. Haque, 2022) In contemporary times, sudden death due to heart attacks has become increasingly prevalent globally, spanning both developed and developing nations. Exploring human inheritance offers valuable insights for advanced techniques development, particularly in identifying individuals predisposed to higher heart attack risks. This study aims to present precise findings regarding the accurate prediction of heart attack likelihood by leveraging previous risk factors and medical histories. Employing machine learning algorithms enables the identification of hidden patterns and tendencies within data, surpassing human predictive capabilities. Consequently, this facilitates personalized and proactive measures for heart attack prevention through early detection strategies.

1. **Introduction:**

The heart is among the most vital organs in the human body, requiring meticulous care, especially for individuals aged sixty and above. Mackay, J., & Mensah, G. A. (2004) According to the World Health Organization (WHO, 2020), an estimated 17.9 million lives are lost annually due to cardiovascular diseases, with over four-fifths of these deaths attributed to heart attacks and strokes occurring prematurely. Researchers are increasingly turning to machine learning classifiers to explore pertinent information. Numerous academic studies have delved into predicting heart disease, spanning both academic and medical domains, utilizing various methodologies and data mining techniques. The authors seek to offer a comprehensive review of research focusing on machine learning applications in the realm of heart conditions. They propose a dataset comprising essential samples and data crucial for the development of effective heart disease prediction methods. Rigorous preprocessing of the dataset is essential to ensure its suitability for machine learning algorithms, ultimately leading to advanced predictive capabilities. Recent advancements include the utilization of various machine learning algorithms to forecast heart attack tendencies, with research addressing critical aspects such as emergency readmission prediction and the development of Stacking Ensemble Learner (SEL) methodologies.

1. **Problem and Data Set Description:**

The dataset utilized in this study was sourced from the Kaggle website, consisting of 303 patient records, and encompassing 14 variables as outlined below. This dataset contains vital medical information regarding patients, indicating the likelihood of experiencing a heart attack. Comprising over 4,000 records and 15 attributes, it offers comprehensive insights into patients' profiles. The variables encompass various demographic, behavioral, and medical risk factors, collectively contributing to a holistic understanding of heart attack susceptibility.

Demographic:

1. Sex: male or female(Nominal)
2. Age: Age of the patient;(Continuous - Although the recorded ages have been truncated to whole numbers, the concept of age is continuous) Behavioural
3. Current Smoker: whether the patient is a current smoker (Nominal)
4. Cigs Per Day: the number of cigarettes that the person smoked on average in one day.(can be considered continuous as one can have any number of cigarettes, even half a cigarette.) Medical( history)
5. BP Meds: whether the patient was on blood pressure medication (Nominal)
6. Prevalent Stroke: whether the patient had previously had a stroke (Nominal)
7. Prevalent Hyp: whether the patient was hypertensive (Nominal)
8. Diabetes: whether the patient had diabetes (Nominal) Medical(current)
9. Tot Chol: total cholesterol level (Continuous)
10. Sys BP: systolic blood pressure (Continuous)
11. Dia BP: diastolic blood pressure (Continuous)
12. BMI: Body Mass Index (Continuous)
13. Heart Rate: heart rate (Continuous - In medical research, variables such as heart rate though in fact discrete, yet are considered continuous because of large number of possible values.)
14. Glucose: glucose level (Continuous) Predict variable (desired target)
15. 10-year risk of coronary heart disease CHD (binary: “1”, means “Yes”, “0” means “No”)

**III. CLASSIFICATION TECHNIQUES**

**Support Vector Classification (SVC):** Is a supervised learning algorithm renowned for enhancing machine learning adaptability by mitigating structured risk, extensively applied in classification tasks. SVC operates as a two-class model aiming to maximize the margin, thereby transforming the problem into a convex quadratic programming solution. Assuming a dataset with N entries, data is organized into input-output pairs, expressed as (xi, yi), where i = 1, 2, ..., N. Here, xi represents the input vector incorporating environmental and individual factors, while yi denotes the output variable reflecting the individual's thermal sensation in the environment. Specifically, yi = 1 indicates a positive class, representing one type of thermal sensation, while yi = -1 signifies a negative class, representing another type of thermal sensation.

**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 predictivemodel 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.

**Naive Bayes:** 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:



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

IV Experimental setup:

From our research, there were no missing values, I carried out normalization of my dataset using Standard Scaler( ). With instances of 303 the data was split into train and test set of 80% and 20% respectively. All the 13 features were used. After training, the dataset was further processed through testing to classify patients tendency of heart attack or not. With, the aid of python performance was evaluated with metrics such as accuracy and classification reports. Check appendix for details

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:

M. Rizwan, S. Arshad, H. Aijaz, R. A. Khan and M. Z. U. Haque, "Heart Attack Prediction using Machine Learning Approach," 2022 Third International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT), Karachi, Pakistan, 2022, pp. 1-8

Mackay, J., & Mensah, G. A. (2004). The atlas of heart disease and stroke. World Health Organization.

[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.