J. Electrical Systems 20-5s (2024): 1428-1436

[1] Romeo Jousef A. Laxamana
[2] Joan Marie Vale

Heart Attack Prediction using Machine Learning Algorithms

Abstract: - One of the most crucial parts of the human body is the heart. When the heart's blood supply is cut off, a heart attack happens. The most frequent cause of a blockage is the buildup of fats, cholesterol, and other substances inside the coronary arteries that provide blood to the heart (which eventually results in plaque growth). This study sought to identify which anthropometric characteristics had a high likelihood of having an impact on a person having a heart attack in order to design a program for heart attack analysis using machine learning algorithms. The researchers were able to acquire data from a variety of sources to identify the variables that may be used to create the output in order to fulfill the study's objectives. The existing heart attack analysis was reviewed by the researchers. They discovered that the researchers' algorithms were hardly noteworthy. Another ongoing study uses a variety of data mining and machine learning approaches to analyze vast volumes of patient data in an effort to predict heart attacks before they happen, ultimately helping professionals and the healthcare industry. The software for Heart Attack Prediction was created by taking into account various data from previous studies.
Keywords: Anthropometric factors, heart attack, Machine learning, prediction

1. INTRODUCTION

One of the most important organs in the human body is the heart. Every time the heart's blood flow is interrupted, a cardiac problem develops. Plaques, which are the most common cause of a blockage, are formed when cholesterol, fat, and other substances assemble in the blood channels that provide blood to the heart. When a plaque ruptures, a clot may form that could impede blood flow. The heart muscle may be harmed or develop malfunction in specific places if blood flow is disrupted. A myocardial infarction (MI), commonly referred to as a "heart attack," occurs when the blood supply to a specific area of the myocardium is reduced or interrupted. Myocardial infarctions can be deadly or "hidden," leading to hemodynamic deterioration and sudden death. The largest leading cause of death in the world and the most frequent cause of myocardial infarction is coronary artery disease (CAD). The myocardium loses oxygen when one coronary artery is blocked. Myocardial loss and necrosis (cell death) can occur when the myocardium is deprived of oxygen for an extended period of time. Patients frequently describe tightness in their chest or pain that spreads to their neck, jaw, shoulder, or arm. In addition to the history and physical examination, myocardial ischemia may also be accompanied by altered ECG patterns and elevated biochemical markers, such as cardiac troponins. Machine learning, which is based on deliberately preparing and testing with the aid of Python and Python libraries, is one of the capable, talented, and effective improvements. Gains in preparation skills for knowledge and information with a clear structure. Testing on various demands should be done in light of this planning and in accordance with key calculations. Machine learning is a compelling breakthrough that can be used for training and testing. It fits in with artificial intelligence (AI). One of the branches of artificial intelligence is AI. It should be possible to use AI innovation to accomplish tasks that are currently carried out by humans using their insight. Machine Learning technology is equipped with a variety of information-using cycles in order to enhance human insight features. The common marvel is used as an ML representation. This forecast must be completed using AI computations and Python libraries. This awareness includes elements of organics like sex, blood pressure, cholesterol, and chest pain. With the use of these elements, six calculations—ML-like Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbors, and Logistic Regression—are used to forecast the results of the research and determine the optimal method based on the Disarray lattice. Given the signs of a heart attack today, many people struggle to determine if they are experiencing one. Due to a lack of access to medical facilities and experts, it could take a while for them to be diagnosed. Patients are often prevented from seeking quick advice due to high healthcare costs. The researchers are interested in finding a solution to this issue. A technique to estimate one's risk of suffering a heart attack was what the researchers sought to develop. They developed an algorithm that calculates the likelihood of a heart attack using variables related to the condition that causes them, coronary artery disease. The goal of the research is to advise people about the likelihood that a heart attack will start based on important parameters. The researchers envisaged the creation of heart attack analysis utilizing machine learning techniques.

[1] Department of Research Development and Extension Services, Faculty of College of Industrial Technology, Batangas State University- The National Engineering University, Balayan, Batangas Philippines,
[2] Department of Computer Studies and Multimedia Arts, Faculty of Computer Science, Far Eastern University Alabang, Muntinlupa City, Philippines, romeojousef.laxamana@g.batstate-u.edu.ph.

2. METHODOLOGY

1. Machine Learning Algorithms

1.1 Random Forest Algorithm

A well-liked algorithm built on the idea of ensemble learning is Random Forest. By merging many learning models, it enhances the outcome of challenging situations. To get more accurate and reliable results, the approach builds numerous decision trees before merging them. The outcome is more precise the more trees there are in the forest. As the level of unpredictability rises, the random forest algorithm creates a forest in the shape of a collection of decision trees. The approach boosts variation and produces a better model by looking for the best features among a random group of features when splitting a node. As a result, only a random subset of the features is checked when splitting a node.

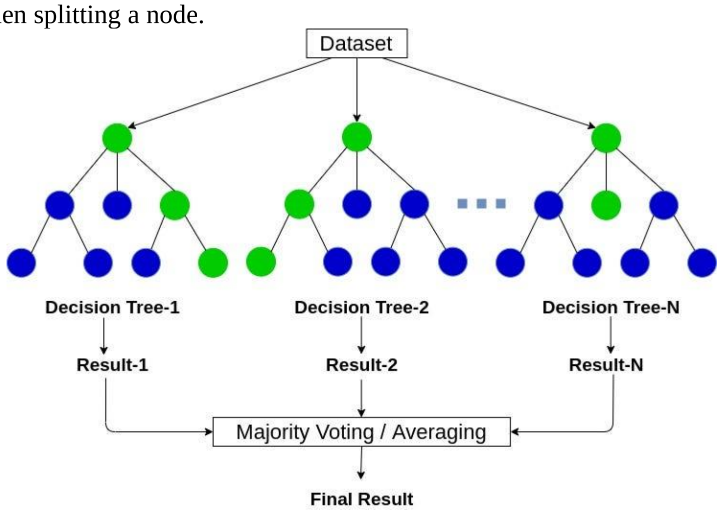

Fig. 1. Random forest is a famous and easy-to-use machine learning algorithm based on ensemble learning (a process of combining multiple classifiers to form an effective model). From Random Forests in Machine Learning: A Detailed Explanation by Saumya Awasthi, 2020.

1.2 Gradient Boosting Algorithm

One of the most potent algorithms in the field of machine learning is the gradient boosting method. As we are all aware, there are two different kinds of machine learning algorithm flaws: those brought on by human mistake and those brought on by machine learning algorithms. Bias error and variance error are the two different types of errors. One of the boosting techniques used to reduce the model's bias error is gradient boosting.

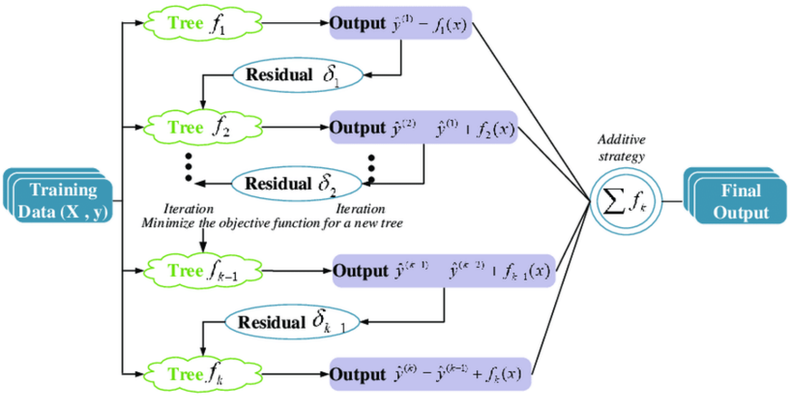

Fig. 2. The structure of extreme gradient boosting. From Simultaneous Determination of Metal Ions in Zinc Sulfate Solution Using UV–Vis Spectrometry and SPSE-XGBoost Method p. 6. Copyright 2020 by Fei Cheng.

1.3 K-Nearest Neighbors (KNN)

A straightforward supervised machine learning method that may be applied to classification and regression issues is the K-Nearest Neighbors algorithm. A supervised machine learning technique trains a function to produce an appropriate output when given additional unlabeled data using labeled input data. We repeatedly apply the K-Nearest Neighbors technique with various values of K to identify the one that reduces errors while maintaining the system's capacity to make accurate predictions when faced with fresh data (Harrison, 29 2018).

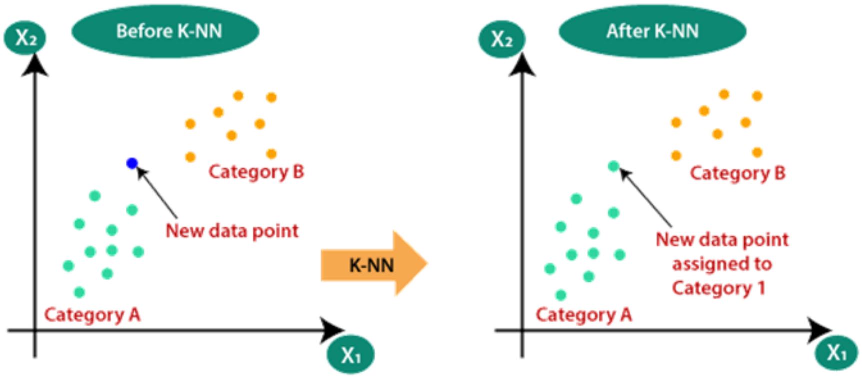

Fig. 3. K-Nearest Neighbours (KNN) Structure. From Let's get to know what the K-Nearest Neighbors (KNN) algorithm is. Copyright 2021 by Trivusi

1.4 Logistic Regression

As a machine learning technique for classification issues, logistic regression is a predictive analytic algorithm built on the idea of probability. A logistic regression model is similar to a linear regression model but uses a more complex cost function, called the sigmoid function or "logistic function," as opposed to a linear function. The cost function is often restricted by the logistic regression hypothesis to the range of 0 and 1. Because they can have a value of greater than 1 or less than 0, which the logistic regression hypothesis claims is not conceivable, linear functions cannot adequately explain it.

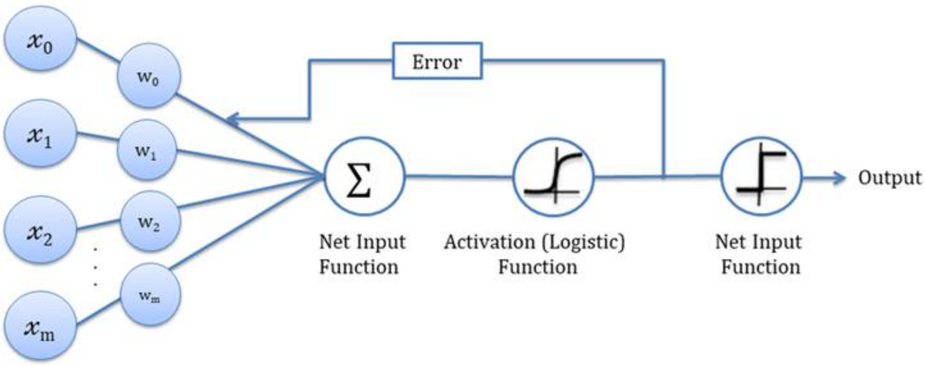

Fig. 4. Schematic diagram for logistic regression classification. From Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques Volume 14, no. 7. by Sabir Hossain, 2019

2. Evaluation Metrics

2.1 Accuracy Rate

A true positive is an outcome for which the model accurately predicted the positive class and is used to determine the accuracy of the model. A true negative is a result for which the model correctly predicts the negative class, much like a true positive. When the model forecasts the positive class incorrectly, a false positive result. False negative results occur when the model predicts the negative class inaccurately.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

2.2 Recall

Recall, the second factor in the F1 Score, can also be used as a standalone machine-learning metric. The recall calculation is shown below:

$$Recall = \frac{TP}{(TP + FN)}$$

2.3 Precision

Precision is the percentage of positively identified instances or samples that were accurately categorized. The precision can therefore be determined by applying the following formula:

$$\textbf{\textit{Precision}} = \frac{TP}{(TP + FP)}$$

2.4 F1-Score

Precision and recall are averaged to produce the F1 score. Since they are both rates, choosing the harmonic mean makes sense. The F1 score formula is as follows:

$$\textbf{\textit{F1-Score}} = 2 * \frac{Precision * Recall}{\text{Precision} + Recall}$$

3. RESULTS

3.1.1 Correlation Matrix of Factors

The correlation matrix is displayed in Figure 5. The red color is 1.0, which denotes a potential heart attack. Now determine which factor (the red color) is closest to 1.0. Cp was 0.433798, age was 0.232219, resting ecg was 0.137230, hypertension was 0.90183, smoking_status was 0.17880, fbs was 0.28046, and sex was 0.280937. It implies that there is a greater likelihood that someone who experiences chest pain will experience a heart attack.

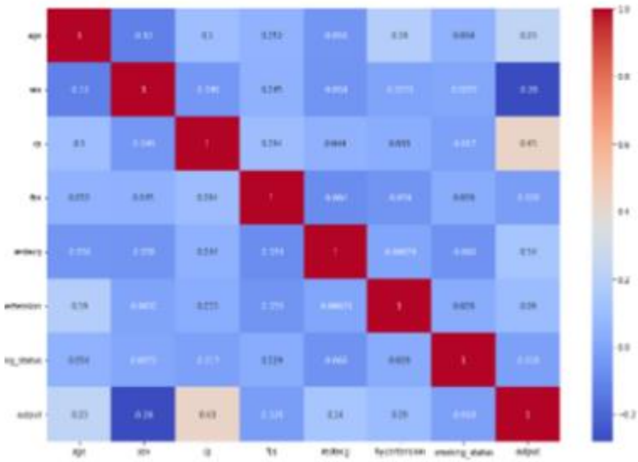

Fig. 5. Correlation Matrix of Anthropometric Factors based on the evaluation results of the authors of this study.

3.1.2 Analysis of Heart Attack Possibility

The potential outcome of a heart attack is depicted in Figure 7. The most significant contributing factor to why someone has a heart attack, according to the graph, is chest pain. According to the medical expert's clinical experience, chest discomfort—especially pain that is anginal in nature—is the main contributing cause to heart attacks. It might be the most significant factor in determining myocardial infarction risk. This symptom is typically what causes patients to consult with medical facilities, which confirms the diagnosis of a heart attack.

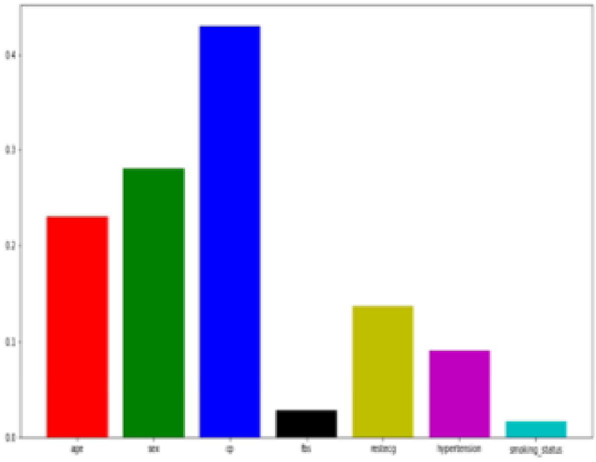

Fig. 6. Heart Attack Possibility in terms of Percentage based on the data preprocessing result of the authors of this study

3.1.3 Validation of Medical Expert

With the aid of medical professionals, Table 1 below shows the Testing and Validation of Heart Attack Analysis. To determine if a patient has a heart attack risk or not, the researchers tested their ability to predict a heart attack by entering data from anthropometric parameters.

Table 1. The validation of medical experts on the prediction results

Trial No.	Input Values for Prediction	Results	Validation of Medical Expert
1	Age = 56 Sex = 1 (male) Cp = 0 (Asymptomatic) Fbs = 0 (non-elevated) Restecg = 2(ST-T abnormality) Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)	No Heart Attack Risk	No Heart Attack Risk
2	Age = 75 Sex = 0 (Female) Cp = 2 (Atypical anginal pain) Fbs = 1 (elevated) Restecg = 2(ST-T abnormality) Hypertension = 1 (hypertensive) Smoking status = 1 (smoker)	Heart Attack Risk	Heart Attack Risk

3	Age = 57 Sex = 1 (male) Cp = 1 (Atypical angina pain) Fbs = 1 (elevated) Restecg = 1(normal) Hypertension = 1 (hypertensive) Smoking status = 1 (smoker)	Heart Attack Risk	Heart Attack Risk
4	Age = 39 Sex = 0 (Female) Cp = 0 (Asymptomatic) Fbs = 0 (non-elevated) Restecg = 0 (hypertrophy) Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)	No Heart Attack Risk	No Heart Attack Risk
5	Age = 80 Sex = 1 (male) Cp = 1 (typical angina) Fbs = 1 (elevated) Restecg = 1(normal) Hypertension = 1 (hypertensive) Smoking status = 0 (non-smoker)	Heart Attack Risk	Heart Attack Risk
6	Age = 41 Sex = 0 (Female) Cp = 0 (Asymptomatic) Fbs = 0 (non-elevated) Restecg = 1(normal) Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)	No Heart Attack Risk	No Heart Attack Risk
7	Age = 58 Sex = 0 (Female) Cp = 1 (Typical Angina) Fbs = 0 (non-elevated) Restecg = 1 (normal)	Heart Attack Risk	Heart Attack Risk

	Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)		
8	Age = 40 Sex = 0 (Female) Cp = 0 (Asymptomatic) Fbs = 0 (non-elevated) Restecg = 1(normal) Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)	Heart Attack Risk	No Heart Attack Risk
9	Age = 37 Sex = 1 (male) Cp = 0 (Asymptomatic) Fbs = 0 (non-elevated) Restecg = 1(normal) Hypertension = 0 (non-hypertensive) Smoking status = 0 (non-smoker)	Heart Attack Risk	No Heart Attack Risk
10	Age = 62 Sex = 0 (Female) Cp = 3 (non-anginal pain) Fbs = 1 (elevated) Restecg = 2(ST-T abnormality) Hypertension = 1 (hypertensive) Smoking status = 0 (non-smoker)	Heart Attack Risk	Heart Attack Risk

4. CONCLUSION

The goal of this work was to develop a tool for heart attack analysis that assesses which anthropometric traits have a high likelihood of influencing a person having a heart attack using machine learning algorithms. The results of this study can help people understand their risk of having a heart attack. The researchers were able to gather data from a variety of sources to identify the variables that may be used to produce the outcome in order to fulfill the study's objectives. They looked at the most recent studies on cardiac attacks. They found that the algorithms of the researchers were not very high. The study's flow was reviewed, with the researchers' exclusive attention going to the implementation of outlier identification, outlier treatment, training models, and model selection. Another current study uses a variety of machine learning and data mining approaches to analyze vast volumes of cardiovascular patient data in order to predict heart attacks before they happen, which is advantageous to the healthcare industry and experts. In order to create the Heart Attack Analysis tool, a number of pieces of data from recent studies were analyzed. To enhance the program of heart attack analysis and obtain better results, the researchers also sought professional advice. The study's construction was based on the knowledge and data that were gathered. Using this information, the researchers were able to produce the output's desired functionality. The team used Google Collab to build the program for cardiac attack analysis. The advocates' preferred programming language was Python. After developing a program for heart attack analysis and carrying out a number of experiments and validations with the assistance of medical professionals, the researchers determined the factors that can influence a potential heart attack. They concentrated on the heart attack outcomes to ensure that they matched their understanding of the potential risk factors for heart attacks. The algorithm utilized in the study was also put to test by the researchers. A specialist was asked to fill the program's requirements so that it could function effectively when the program for heart attack analysis was delivered.

REFERENCES

[1] Shruti Kothadia (2021). Classification Algorithms in Python – Heart Attack Prediction and Analysis.

[2] T. Obasi and M. Omair Shafiq (2019). "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2393-2402.

[3] Suraj Kumar Gupta, Aditya Shrivastava, Satya Prakash Upadhyay, & Pawan Kumar Chaurasia. (2021). A Machine Learning Approach for Heart Attack Prediction. International Journal of Engineering and Advanced Technology (IJEAT), 10(6), 124–134.

[4] Tripathi, Ashish & Singh, Arun & Singh, Krishna Kant & Choudhary, Pushpa & Vashist, Prem. (2021). Machine learning architecture and framework.

[5] Priya, P., Aluru, N.R. Accelerated design and discovery of perovskites with high conductivity for energy applications
through machine learning. (2021). *npj Comput Mater* **7**, 90

[6] Millett ERC, Peters SAE, Woodward M (2018). Sex differences in risk factors for myocardial infarction: cohort study of UK Biobank participants.

[7] Jarvis S (2021) Electrocardiogram 2: interpretation and signs of heart disease. *Nursing Times* [online]; 117: 7, 51-55

[8] Bass, Christopher. (2002). ABC of psychological medicine: Chest pain. British Medical Journal - BMJ. 325. 588-591.

[9] Chadwick Jayaraj, Joshua, Karapet Davatyan, S.S. Subramanian, and Jemmi Priya (2019). 'Epidemiology of Myocardial
Infarction'. Myocardial Infarction. IntechOpen.

[10] Pol, Tymon & Held, Claes & Westerbergh, Johan & Lindbäck, Johan & Alexander, John & Alings, Marco & Erol, Cetin
& Goto, Shinya & Halvorsen, Sigrun & Huber, Kurt & Hanna, Michael & Lopes, Renato & Ruzyllo, Witold & Granger,
Christopher & Hijazi, Ziad. (2018). Dyslipidemia and Risk of Cardiovascular Events in Patients With Atrial Fibrillation
Treated With Oral Anticoagulation Therapy: Insights From the ARISTOTLE (Apixaban for Reduction in Stroke and
Other Thromboembolic Events in Atrial Fibrillation) Trial. Journal of the American Heart Association. 7

[11] Gallucci, Giuseppina & Tartarone, Alfredo & Lerose, Rosa & Lalinga, Anna & Capobianco, Alba. (2020). Cardiovascular risk of smoking and benefits of smoking cessation. Journal of Thoracic Disease. 12. 3866-3876.

[12] Ginghina, Carmen & Ungureanu, Catalina & Vladaia, Aurora & Popescu, Bogdan Alexandru & Jurcut, Ruxandra. (2009). The electrocardiographic profile of patients with angina pectoris. Journal of medicine and life. 2. 80-91.

[13] Naveen (2023). "What Is Precision, Recall, Accuracy and F1-Score?" *Nomidl*, July.

[14] Patel, Nikunj, PrinceKumar Patel, and Nehal Patel. "Heart attack detection and heart rate monitoring using IoT." *International Journal of Innovations and Advancements in Computer Science, IJIACS* 7, no. 4 (2018): 612-615.

[15] Keya, Maria & Shamsojjaman, Muhammad & Hossain, Faruq & Akter, Farzana & Islam, Fakrul & Emon, Minhaz. (2021). Measuring the Heart Attack Possibility using Different Types of Machine Learning Algorithms.

[16] Sharma, Ochin. (2023). Prediction and Analysis of Heart Attack using Various Machine Learning Algorithms. 786-790.

[17] Bokshi, Lutfar & Al Banna, Md & Ghosh, Tapotosh & Nahian, Md Jaber & Kaiser, M. Shamim. (2022). Investigation on Heart Attack Prediction Based on the Different Machine Learning Approaches.
[18] Anas, Rakibul. (2023). Heart Disease Prediction Using Machine Learning Techniques.
[19] Rizwan, Muhammad & Arshad, Sadia & Aijaz, Hafsa & Khan, Rizwan & Haque, Dr. (2022). Heart Attack Prediction using Machine Learning Approach.
[20] Salhi, Dhai Eddine & Tari, Abdelkamel & Kechadi, Tahar. (2021). Using Machine Learning for Heart Disease Prediction.