International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Heart Attack Risk Prediction Using XG Boost Algorithm

P. Rajapandian[1], S. Kiruthisha[2]

[1] (Associate Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College,Pondicherry-605 107.
Email: rajapandian.mca@smvec.ac.in)

[2](P.G Student, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry-605 107.
Email: kirithishashanmugam0@gmail.com)

----------**********----------

Abstract:

Heart Attack Risk Predictive Simulator is a web-based health solution for evaluating your potential risk for a heart attack that utilizes machine learning methods. The Heart Attack Risk Predictive Simulator uses machine learning techniques by being fed medical history information from a patient, including things like age and cholesterol levels (i.e., risk factors) and generates a prediction of your chances of having a heart attack in the next 12 months. The prediction is made quickly with the most accurate measurements available, and we will accomplish this with the XG Boost algorithm. The simulator will be built using a technology stack containing both Django and MySQL. The Heart Attack Risk Predictive Simulator will allow users to enter their own personal medical information, provide predictive data concerning their chance of experiencing a heart attack; recommend preventative health services based on previous data entered, and generate a predictive health (pdf) report based on the predictive data for later review. The overall goal of creating the simulator is to improve the prevention of disease through the ability to accurately predict a person's future risk of experiencing a heart attack.

Keywords — Heart Attack Risk Prediction, XG Boost Algorithm, Machine Learning, Healthcare Analytics, Disease Prediction, Preventive Healthcare, Medical Data Analysis, Clinical Decision Support System

----------**********----------

I. INTRODUCTION

Every year, Coronary Artery Disease (CAD) leads to more than 1.5 million deaths worldwide, making it one of the top three causes of mortality globally. Numerous people die from this disease due to a combination of diseases and problems that are associated with lifestyle changes as well as modern living. The leading contributing factors for this disease are due to our diets (poorly made and poor nutritional content) and lack of activity (exercise); however, the greatest reason people develop CAD today is due to their overall lifestyle and many other reasons such as smoking (to name one). This has increased the number of people developing CAD among many different age groups, thereby amplifying the global risk of cardiovascular disease and the chances of having a fatal heart attack. As a result of this concern, early identification of cardiovascular disease is essential because if treated

early enough, preventative treatment can help reduce the likelihood of someone developing a serious health condition or dying due to CAD. The medical profession has been able to develop several ways to evaluate a patient to identify cardiovascular conditions. However, this process usually requires a great deal of testing (diagnostic testing), evaluation of the patient's health (medically), as well as consulting with multiple medical professionals, which creates a cumbersome and time-consuming method for diagnosing cardiovascular disease..

Technological advances, specifically in the tech field, have made predictive analytics one of the major ways that ML is being utilized in healthcare; this is primarily due to the continuing advances in many of the algorithms that are created for ML. Predictive analytics is a very accurate and quick method for predicting disease occurrences. This project describes a heart attack risk prediction system that has been developed through the use of an XG Boost algorithm based on patient health-related data (age, cholesterol level, blood pressure, heart rate, type of chest pain). The end product is a web-based application that was created using the Django framework, and that allows users to easily input their relevant health-related data into a website to receive an instant estimate of their risk for a heart attack and personalized suggestions for leading a healthier life or receiving healthcare, nutrition and exercise.

II. LITERATURE SURVEY

1. S. Patel and R. Kumar (2021) conducted an analysis of the use of machine learning methodologies to identify patients at risk for developing heart disease as an early intervention strategy through the prediction of heart disease diagnosis through algorithms of Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN). Their results showed that machine learning models have the potential to help healthcare professionals improve both the speed of their predictions and make better decisions based on the information available from the patient's medical history. They further indicate that traditional predictive models are only moderately accurate, as well as being less than completely reliable because the size of the health dataset in question is significant or large.

2. A. Sharma & P. Verma, 2020, Heart Disease Prediction Using ML Algorithms. This article discusses various types of ML algorithms that can be applied to the prediction of heart disease, using the medical records of patients. The authors primarily compare Logistic Regression, Random Forest, and Support Vector Machine (SVM) to determine which method is best for use in predictive technology associated with heart disease diagnosis. The authors believe that the application of ML will help reduce the amount of manual effort involved in diagnosing heart disease and there is greater potential for the early identification of heart disease than through traditional methods. However, the authors note that the accuracy of the predictive algorithm and the ability of the method to efficiently handle complex medical data are both limitations on the performance of the system.

3. In the paper, "Predicting Heart Disease with Data Mining Methods" (Singh, K. & Gupta, M., 2019), the authors examine the use of data mining and machine learning to predict risk for heart disease based on patient attributes such as age, cholesterol level, blood pressure and type of chest pain. Specifically, the authors indicate that the KNN and Decision Tree algorithms can assist medical prediction systems but also note that an algorithm may produce varying levels of prediction depending on the quality of the data and the selected algorithm. Based upon this, the authors proposed that there should be development of the different types of machine learning algorithms to improve healthcare prediction accuracy and efficiency.

III . PROBLEM STAEMENT

Heart disease is now among the leading causes of death worldwide, with identifying risk for heart attack early on being one of the most difficult tasks. Traditional diagnostic methods often require

multiple medical tests, experts and take a long time, potentially delaying preventive treatment. Additionally, current predictive systems performing simple machine learning algorithms sometimes lack accuracy and efficiency when working with health-care data. As such, we need an intelligent and dependable approach to predicting heart attack risk using patient medical data, with timely feedback to permit early intervention and improve healthcare decision making

IV. EXISTING SYSTEM

Heart attack prediction systems currently in place mainly consist of machine learning algorithms (such as Logistic Regression, Random Forest or KNN) to assess the likelihood that someone will suffer from heart disease.

These systems develop a working estimate of a patient's likelihood of having a heart attack by looking at their medical records, including their age, blood pressure, cholesterol level, type of chest pain, and heart rate.

The current systems will also allow for early detection of an individual disease as well as give them an overall prognosis of their health.

However, the vast majority of traditional heart attack prediction systems are only able to achieve between 75 - 80 percent accurate predictions which of course means there is a high level of inaccuracy associated with the results of these systems.

Moreover, many heart attack prediction systems do not perform very well when evaluating medical databases and thus provide less efficient performance and lower prediction quality than was intended when designed.

Also, very few heart attack prediction systems provide tailored information and preventative care for users after a prediction is made.

V. PROPOSED SYSTEM

At present, this system has been developed to use XG Boost as a prediction algorithm for heart disease, which will provide more accurate and efficient predictions compared to previous methods of prediction.

In order to produce accurate heart attack predictions, the following patient medical information will be inputted into the XG Boost algorithm: age, blood pressure/ B.P., cholesterol level, type of chest pain, heart rate and other health-related variables.

By using XG Boost for analyzing health care data, there is an increased level of predicted performance and accuracy compared to conventional machine learning methods.

The development of this system was accomplished by utilizing the Django framework; therefore, users are able to input their medical information and receive an instantaneous prediction of their risk for heart attack by means of a user-friendly web interface.

Also as a part of the prediction that is given to users, the system will also provide them with health care consultative recommendations .In addition to the prediction and health care recommendations, the users of the system will receive a PDF report that can be downloaded and saved for future reference to aid in potential medical treatment.The outcome results of this proposed system is to improve heart attack prediction accuracy, efficiency, and preventative health care resources for users.

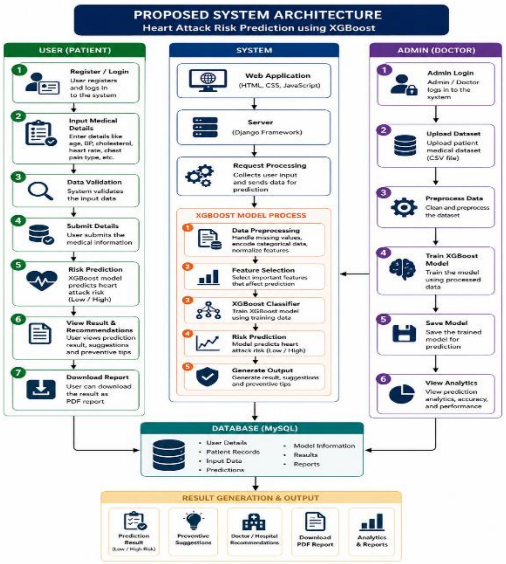

Figure 1: Workflow Of Proposed System

VI. METHODOLOGY

The methodology of the proposed system follows a step-by-step process for conducting examinations and evaluating student responses efficiently**.**

Step 1: Collection of Data
The system gathers information about patients' health from a healthcare database. This data includes key parameters for assessing the risk of a heart attack including but not limited to age, blood pressure, blood cholesterol level, heart rate, chest pain type, glucose level, as well as other health parameters associated with increased risk for heart attack.

Step 2: Pre-Processing of the Data
After the system has collected data, the database is pre-processed to enhance the quality of the data, as well as enhance prediction accuracy for the subsequent use of the data for training and testing purposes. This process includes replacing missing values, removing or correcting unnecessary and inconsistent data.

Step 3: Data Splitting
After the cleaning of datas, datasets are divided for testing and training . The machine learning model will be trained using the training data, while the machine learning model performance and the prediction accuracy will be evaluated against the testing data.

Step 4: Model Training with the XG Boost Algorithm
The XG Boost algorithm will be applied to train the model with the patient healthcare data. The model can learn the patterns associated with the medical attributes of each patient and thus will allow for improved prediction accuracy in identifying the risk of heart attack for each patient

Stage 5: Predictions are made.
Inputting the medical information into the portal on-line the data will be processed by the trained XG Boost. The output will be low or high risk of having a heart attack.

Stage 6: Recommendations are generated.
Using the output from the prediction, there will be generation of recommended lifestyle changes such as diet, exercise, preventive measures and consultation with a physician, robust enough to allow for effective long-term health management. The predictions and recommendations will be saved and a PDF will be created allowing the user to download and use for future reference regarding their health care.

VII. SYSTEM ARCHITECTURE

The Heart Attack Risk Prediction System has a web-based structure that allows users to interact with the predictive model, backend services, and database. The system collects information about a patient's health history and uses it to predict his or her risk of having a heart attack and to provide recommendations regarding the appropriate healthcare for the patient. Through the application, users will be able to enter their health history, see the results of their heart attack risk evaluation, and download a report.

The front end of the application has been built with HTML, CSS, and JavaScript, in order to create an effective user interface. The backend of the application has been developed with Python Django in order to manage user authentication, execute predictive algorithms for evaluations, and store data. The data is stored in a MySQL database. The XG Boost algorithm has been implemented in order to analyze the medical history of the patient and to create predictive models for forecasting a patient's risk of having a heart attack.

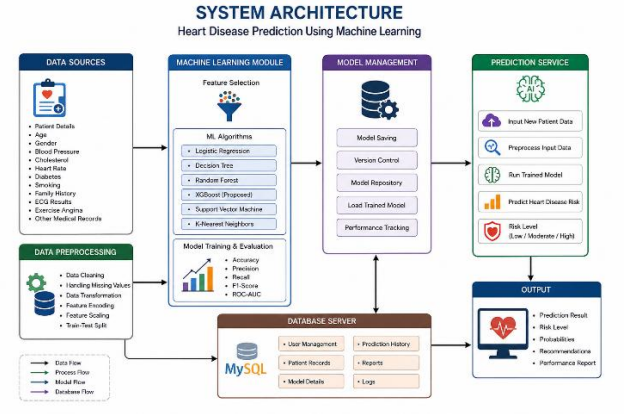

Figure 2 : System Architecture

VIII. MODULES DESCRIPTION

Different modules have been designed to allow a smooth prediction, healthcare recommendations and report generation—each performing the required function of the system**.**

1. User Module

The User Module is designed for users to register/login, enter their medical information in order to obtain their risk prediction for a heart attack. Users will be able to provide health-related information via this system and will receive their prediction results.

Functions:

- User registration and login
- Enter medical data
- View prediction results
- Print PDF reports

2. Prediction Module

The Prediction module is responsible for analysing users' medical data to generate heart attack risk prediction using the XG Boost algorithm for maximum accuracy and efficiency in generating a prediction.

Functions:

- Medical data analysis
- Prediction of heart attack risk
- Determining users' level of risk
- Generation of prediction results

3. Suggestions

The Recommendation Module gives suggestions to help you with your health based on your prediction. You will receive a selection of the following suggestions

- Diet recommendations
- Workout recommendations
- Preventive recommendations
- Referral to see a doctor.

4. Administration module

The tool helps the administrator of the system monitor user activity and all activity related to predictions. The administrator will be able to manage system types and review user's prediction requests and details through the tool’s dashboard.

- Administrator login
- Registered users
- Monitor/Review predictions
- Review of user's history.

5. Report Creation Module

This report creation module produces downloadable PDF reports that contain the output of the prediction and suggestions for the future medical record.

Functions:

- PDF report creation
- Prediction Summary
- Recommendations are stored
- Download report option

IX.WORKING OF XG BOOST ALGORITHM

In this project we are predicting a person's risk of having a heart attack using patient medical history and the XG Boost algorithm. The first step is to gather the user's medical history (including age, blood pressure, cholesterol level, heart rate, chest pain type etc.) and compare it with a previously established set of medical records to determine relationships with history of heart disease.

The XG Boost algorithm creates many different decision trees and enhances each of them by correcting their mistakes in terms of predictions.

Once the XG Boost model is created, it is continuously being trained or updated by the set of medical records resulting in improved predictive capabilities. Based on the user's medical record, the program then predicts whether the patient is at a low or high risk for having a heart attack.

The XG Boost algorithm improves the way it learns from historical data (i.e., data from the past) in order to predict future outcomes and improve its overall predictive performance as part of the training phase.

It uses the relationships between different types of medical attributes and determines whether the patient belongs to either a low or high risk category. It is then able to enhance predictive accuracy by decreasing the chance that a prediction will not happen and creating valid results based on medical data.

The use of XG Boost in this project is due to its therefore better efficiencies and speeds in addition to providing better predictive accuracies over traditional machine learning algorithms. This system can achieve an approximate 90% rate of accuracy in predicting future events for early heart attack risk determination and to provide medically preventative types of health care.

X. RESULTS &CONCLUSION

The proposed Heart Attack Risk Prediction System functioned properly and allowed for the prediction of heart attack risk based on the medical histories of the patients by using the XG Boost algorithm. The system analyzed healthcare parameters (such as age, blood pressure and cholesterol levels, heart rate and chest pain type) to provide prediction results very efficiently. Our model produced nearly 90% prediction accuracy and performed better than older traditional machine learning methods. Based on the prediction results, it generated appropriate healthcare recommendations and provided downloadable PDF reports that could be accessed as a reference for use at future medical appointments. In conclusion, the proposed system significantly supported the early identification of heart attack risk and assisted in improving decision making related to preventative captured to decision making relative to preventative health.

XI. FUTURE ENHANCEMENT

The prediction system for heart attack risk may see additional features (in the future) being added such as doctor consultation recommendations, nearby hospitals, and healthcare-related chat bots, as well as the ability to monitor a user's health remotely in real-time via wearable devices. A mobile app may also be established to allow users to quickly view their prediction results and receive notifications. The addition of these services will allow for the prediction system to be more intelligent, accurate and beneficial for preventative health care management.

XI1. REFERENCES

Patel, S. & Kumar, R.; Machine Learning Approaches for Heart Disease Prediction; International Journal of Healthcare Technology, Volume 12 (3); Pages 145-158; 2021

Sharma, A. & Verma, P.; Heart Disease Prediction Using Machine Learning Algorithms; Journal of Medical Informatics and Healthcare Systems, Volume 10 (2); Pages 78-89; 2020

Singh, K. & Gupta, M.; Prediction Of Heart Disease Using Data Mining Techniques; International Journal Of Computer Applications in Healthcare, Volume 8 (4); Pages 56-67; 2019

Brown, J. & Wilson, L.; Healthcare Prediction Systems Using Machine Learning; Journal of Artificial Intelligence in Medicine, Volume 14 (1); Pages 90-105; 2022

Thomas, M. & George, R.; Early Heart Disease Prediction Using Classification Algorithms; International Journal of Advanced Research in Computer Science, Volume 11 (5); Pages 66-74; 2021