International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Heart Disease Prediction Using Machine Learning Techniques in Python: A Review

Tanmay Deshmukh, Supriya Kharatmol, Professor Nishant Patil
MGM College of Engineering & Technology, Kamothe, Navi Mumbai

Abstract- As the global incidence of heart disease escalates daily, there is an urgent imperative to accurately predict and diagnose these conditions efficiently. Heart illness, also referred to as cardiovascular disease, is a broad category of conditions that affect the heart, including congenital abnormalities, vascular problems, and cardiac arrhythmias. In recent decades, it has emerged as one of the world's top causes of death. Thus, it is imperative to create accurate and trustworthy techniques for early disease detection .Heart illness, also referred to as cardiovascular disease, is a broad category of conditions that affect the heart, including congenital abnormalities, vascular problems, and cardiac arrhythmias

Index Terms- Heart Disease, Machine Learning, Data Science, Data exploration, Heart Rhythm

I. INTRODUCTION

A heart attack, synonymous with acute myocardial infarction (AMI), stands as a grave ailment within the realm of cardiovascular disorders. It manifests when blood flow to the heart muscle is obstructed, resulting in damage to the cardiac tissue. The diagnosis of heart disease is paramount, necessitating a keen understanding of its symptoms, thorough physical examination, and recognition of various indicative signs. Factors contributing to heart disease encompass a spectrum of influences, including elevated cholesterol levels, genetic predispositions, hypertension, sedentary lifestyles, obesity, and tobacco usage. A primary cause of heart attacks lies in the cessation of blood flow through the coronary arteries. As blood flow diminishes, red blood cell counts decline, leading to oxygen deprivation and potential loss of consciousness. Timely identification of symptoms and signs holds paramount importance, potentially averting heart attacks through accurate prediction and proactive intervention.

Heart diseases have emerged as a leading global cause of mortality, claiming approximately 17.7 million lives annually, which accounts for 31% of all global deaths, according to the World Health Organization (WHO). This trend is also prominent in India, where heart-related ailments have become the primary cause of death, claiming 1.7 million lives in 2016 alone. The economic burden is substantial, with estimates suggesting that India lost up to $237 billion between 2005 and 2015 due to cardiovascular diseases. Given the overwhelming volume and potential noise in the health-related data collected by medical organizations worldwide, leveraging machine learning techniques becomes crucial for accurate prediction. Machine learning offers the capability to analyze complex interactions among various risk factors, thus enhancing cardiovascular risk prediction accuracy. The development of a graphical user interface (GUI) provides a user-friendly platform to interact with heart disease prediction systems. This interface simplifies the input process, allowing users to input key information such as age, sex, blood pressure, cholesterol levels, and other health attributes. Additionally, it presents the predicted probability of heart disease occurrence alongside graphical representations, facilitating easy interpretation. The GUI further enhances transparency.

II. LITERATURE SURVEY

Senthil Kumar Mohan et al [1] proposed a half breed machine learning approach to improve the precision of cardiovascular illness forecast, accomplishing a exactness level of 88.7% through the HRFLM demonstrate, which combines arbitrary woodland with direct demonstrate procedures. Different information mining approaches. Various data mining approaches and prediction methods such as KNN, LR, SVM, NN, and vote have picked up notoriety for heart disease identification and prediction.

Ali, Liaqat, et al [2], presented a framework comprising two linear Support Vector Machine (SVM) models, specifically L1 regularized and L2 regularized, with the previous pointed at highlight determination selection by invalidating coefficients of pointless highlights and the last mentioned for illness forecast. They utilized a crossover lattice look calculation to optimize both models based on measurements like exactness, affectability, straightforwardness, Matthews relationship coefficient, ROC chart, and zone beneath the bend,

© 2024 IJSRET

accomplishing noteworthy enhancements over routine SVM models by 3.3%.

Abhay Kishore created a heart assault expectation framework utilizing profound learning strategies, particularly Recurrent Neural Networks, which proved highly effective in predicting heart-related ailments with minimal errors, offering a promising direction for future development in heart attack prediction platforms.

V.V Ramalingam et al proposed heart illness forecast utilizing machine learning methods, highlighting the expanding application of administered learning calculations such as support vector machines, K-nearest neighbour (KNN), Naïve Bayes, Decision Trees (DT), and ensemble models in automating the analysis of medical datasets for heart disease diagnosis.

This paper presents a study of different models based on such calculations and strategies and analyze their execution. Models based on supervised learning algorithms such as Support Vector Machines, K- nearest neighbours, and ensemble models are found to be very popular among researchers. Strategies to enable the well-being of the industry and the experts in the analysis of heat-related sickness.

"A GUI-based Heart Disease Prediction System Using Machine Learning Procedures" by M. H. Parvez and M. M. Islam.

This study proposed a GUI-based heart disease prediction system using machine learning methods such as decision trees, random forests, and support vector machines. The proposed system achieved an accuracy of 86.67% using the random forest algorithm.

Heart Disease Prediction Using Machine Learning with GUI in Python" by S. S. Rathod and S. D. Thackrey. This research proposed a GUI-based heart disease prediction system using logistic regression, decision trees, and k-nearest neighbour algorithms. The proposed framework accomplished an exactness of 90.50% utilizing the decision tree algorithm.

"Design and Development of GUI Based Heart Disease Prediction System using Machine Learning Algorithms" by P. G. Rao and M. N. Nandish.

This study proposed a GUI-based heart disease prediction system using a decision tree, random forest, and support vector machine algorithms. The proposed system achieved an accuracy of 85.66% using the random forest algorithm.

III. DATASET

We performed a computer recreation on one Dataset. The dataset is the Heart dataset. The dataset is available in the UCI Machine Learning Repository.]

Table 1: Features selected from Dataset.

SR. No.	Attributes Description	Distinct Values of Attributes
i	Age- age of a person	Values 30-70
Ii	Sex- gender of the person(0-Female 1-Male)	0-1
Iii	CP- severity of chest pain the patient is suffering	0, 1, 2, 3
Iv	BP- Patient's BP	Values between 94-200
v	Chol- cholesterol level of the patient	Values between 126 -564
Vi	FBS- fasting blood sugar in the patient	0, 1
Vii	Heartbeats	71 - 202
Viii	Exang- identify if there is exercise-induced angina. If yes=1 or else No=0	0, 1
Ix	Slope- patient's condition during peak exercise. Divided into three segments (up-sloping, Flat, Down sloping)	1, 2, 3
X	Target- It is a class or label column. This dataset has binary classification i.e. two classes(0, 1).	0, 1 0- chances of heart disease is less. 1- chances of heart disease is high.

IV. PROPOSED SYSTEM

The data set consists of the patient's physiological parameters and health records. Then multiple filters are applied to the dataset and pre-processing is done. Extraction of image features is performed we divide the data into two sets: the

© 2024 IJSRET

training data, comprising 80% of the dataset, and the testing data, which makes up the remaining 20%.

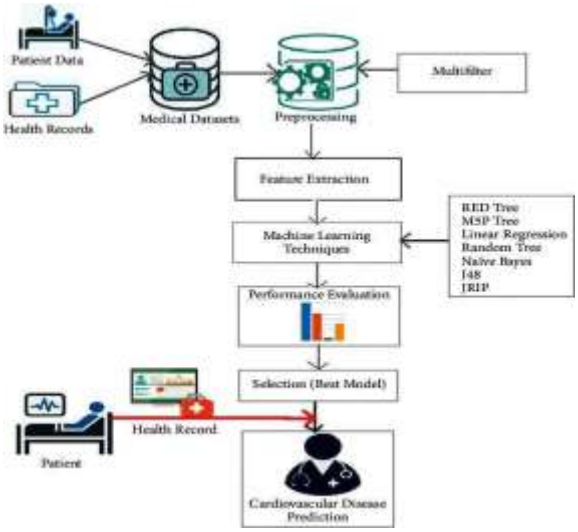


Fig- 1: Block Diagram

Within the training set, known outputs are provided, allowing the model to learn from this information and generalize its predictions to new data instances. We need to apply different machine learning techniques to train the dataset and to calculate the accuracy of the model. After the training of the model then we test the model for the remaining 20% data. In this phase, we give only inputs to model and predict the output. Then compare the predicted output with the actual result and find out parameters such as accuracy, sensitivity, precision, recall, specificity, etc. We have to select the best model which has greater accuracy than other models.

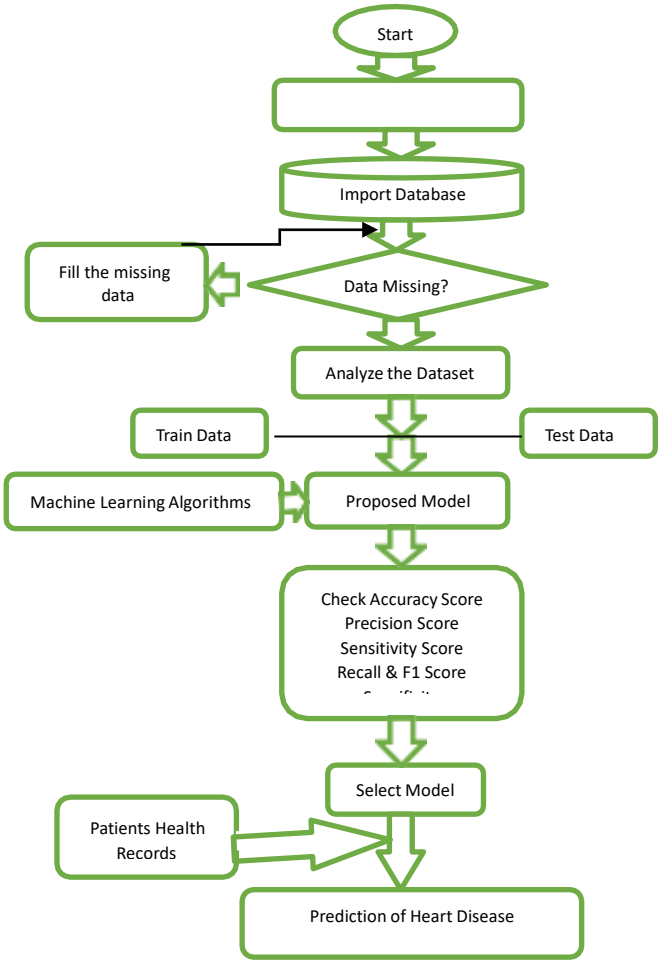
1. GUI Implementation

A graphical user interface (GUI) serves as a digital platform facilitating user interaction with graphical elements like icons, buttons, and menus. Through the GUI, users receive visual representations conveying pertinent information and offering interactive options for actions they can undertake.

In our thesis, we have created a GUI Interface for Heart Disease Prediction Using Machine Learning Techniques in Python. First, we start the process with data collection, in our thesis work we collected a CVD dataset from Cleveland. The dataset comprises 304 patient records, each consisting of 10 features including age, gender, systolic blood pressure, diastolic blood pressure, etc. The target class "cardio" is assigned a value of 0 for healthy patients and 1 for those diagnosed with cardiovascular disease. The objective is to forecast the presence or absence of cardiovascular disease (CVD) based on the results of patient examinations. The second step is importing all necessary Python libraries, by using the Jupiter or Spyder IDE. Here we have used Spyder IDE to implement our code. Here, we considered 80% of the dataset to be the training data and the remaining 20% to be the testing data. Then the Separated training and testing data is an important part of evaluating data models. Then we used 4 different ML techniques to predict heart disease.

In this user-friendly GUI, we just need to enter the patient data and click on the button and we will get the result of whether the person is suffering from heart disease or not.

2. Flowchart



Import Patient Data: The patient data is imported into the system from various sources such as electronic health records, medical devices, or surveys. Import database: The database containing relevant health information is imported into the system.

Data Missing: If there is any missing data, the system prompts you to fill in the missing data with relevant information.

© 2024 IJSRET

Analyze Dataset: The dataset is analyze to understand the distribution of data, and correlations between features, and identify any patterns

The dataset is split into two portions: 80% for training and 20% for testing. The training data is utilized to develop and train the machine learning model, while the testing data is The dataset is split into two portions: 80% for training and 20% for testing. The training data is utilized to develop and train the machine learning model, while the testing data is used to assess the model's performance and accuracy.

The decision tree algorithm is a type of supervised learning method that can address both regression and classification problems. Its goal is to develop a training model that uses simple decision rules learned from prior data (training data) to predict the class or value of a target variable. Apply machine learning algorithms: Various machine learning algorithms such as Logistic Regression, Decision Tree, SVM, KNN, etc. are applied to the dataset to create a predictive model.

The decision tree algorithm is a type of supervised learning method that can address both regression and classification problems. Its goal is to develop a training model that uses simple decision rules learned from prior data (training data) to predict the class or value of a target variable.

4.3 Machine Learning Methods

a) **SVM:** Support vector machines (SVM) come in different types, such as linear and nonlinear classifiers. Typically, two datasets are used with SVM: a training set and a test set. In an ideal scenario, the classes are linearly separable, meaning a line can be found that perfectly divides the two classes. However, multiple lines may achieve this separation. The optimal line, known as the "separating line," is chosen from these options based on maximizing the margin between the classes. This line ensures the most effective classification and generalization performance of the model

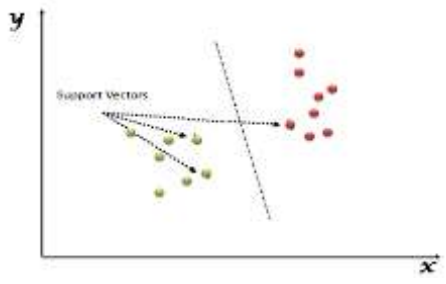


Fig- 2:.a SVM

Decision Tree: The decision tree algorithm is a type of supervised learning method that can address both regression and classification problems. Its goal is to develop a training model that uses simple decision rules learned from prior data (training data) to predict the class or value of a target variable.

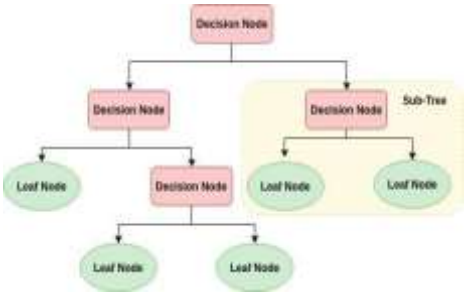


Fig- 3.b Decision Tree

c) **KNN:** The K-nearest neighbours algorithm, commonly known as KNN, is a non-parametric, supervised learning classifier that relies on proximity to classify or predict the grouping of individual data points. Although it can be applied to both regression and classification problems, KNN is most often used for classification, based on the assumption that similar points tend to be located near each other

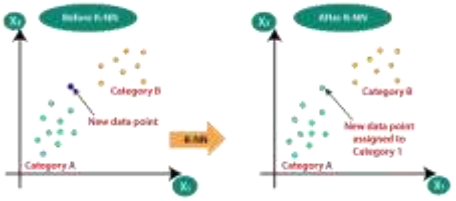


Fig- 4.c SVM

Logistic Regression: Logistic regression is a supervised machine learning algorithm designed for binary classification tasks by predicting the probability of an outcome, event, or observation. The model provides a binary outcome, limited to two possibilities such as yes/no, 0/1, or true/false. Logistic regression examines the relationship between one or more independent variables and classifies data into discrete classes. It is widely used in predictive modelling to estimate the probability of an instance belonging to a specific category.

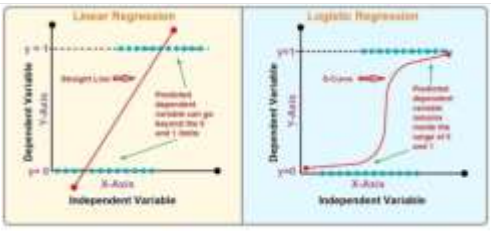


Fig- 5 .d Logistic Regression

© 2024 IJSRET

4. Software Used

Python: A web scraper written in Python is used to collect data for the model. According to Wikipedia, Python's syntax allows programmers to express concepts in fewer lines of code. Python's implementation began in December 1989 by Guido van Rossum at CWI in the Netherlands. Python 2.0 was released on October 16, 2000, while Python 3.0 was released on December 3, 2008.

Python provides a module called 'urllib2' that offers functions to open websites and easily extract information. The web scraper, programmed in Python, is responsible for collecting weather data for the model.

Ms Excel

Microsoft Excel is a spreadsheet application developed by Microsoft for Windows and Mac OS X. It offers features such as calculations, graphing tools, pivot tables, and macro programming. language. The first version of Microsoft Excel was released in 1987. It is a comprehensive spreadsheet tool that supports various file extensions and offers numerous features. Its user-friendly interface is helpful, but in cases where it falls short, there are other alternatives to explore

V. Result

1. To predict the likelihood of heart disease, we need to input the following medical parameters

- **Age**: 63
- **Sex**: Male
- **Chest Pain Type**: Type 1 (typical angina)
- **Resting Blood Pressure**: 145 mm Hg
- **Serum Cholesterol**: 233 mg/dL
- **Fasting Blood Sugar**: High (greater than 120 mg/dL)
- **Resting Electrocardiographic Result**: Indicates possible abnormality (type 2)
- **Maximum Heart Rate Achieved**: 150 beats per minute
- **Exercise-Induced Angina**: No
- **ST Depression Induced by Exercise Relative to Rest**: 2.3 mm
- **Slope of the Peak Exercise ST Segment**: Down sloping (type 3)
- **Number of Major Vessels Coloured by Fluoroscopy**: None (0)
- **Thalassemia**: Fixed defect (type 6)

Fig. 5.1 Graphical User Interface for Heart Disease Prediction System (Person does not have Heart Disease)

Result: The results of different machine learning models for heart disease prediction, based on various performance metrics **Logistic Regression**:

- **Accuracy**: 0.80
- **Precision**: 0.85
- **Recall**: 0.65
- **Specificity**: 0.91

Support Vector Machine (SVM):

- **Accuracy**: 0.82
- **Precision**: 0.86
- **Recall**: 0.69
- **Specificity**: 0.91
- **K-Nearest Neighbours (KNN)**:
  - **Accuracy**: 0.70
  - **Precision**: 0.67
  - **Recall**: 0.62
  - **Specificity**: 0.77
- **Decision Tree**:
  - **Accuracy**: 0.77
  - **Precision**: 0.73
  - **Recall**: 0.73
  - **Specificity**: 0.80

Firstly we have to enter the medical parameters that are required for heart disease prediction which include: Age:60, Sex:1, Chest Pain type:2, Resting Blood Pressure:130, Serum Cholesterol:270, Fasting Blood Sugar:1, Resting electrocardiographic result:2, Maximum heart rate achieved:140, Exercise-induced angina:1, ST depression induced by exercise relative to rest:2.3, Slope of the peak exercise ST segment:2, Number of major vessels coloured by fluoroscopy:3, Thalassemia:6

© 2024 IJSRET

Fig. 5.2 Graphical User Interface for Heart Disease Prediction System (Person having Heart Disease)

Result: The person has heart disease.

- Values we are getting using Logistic Regression: Accuracy: 0.80, Precision: 0.85, Recall: 0.65, Specificity: 0.91.
- Values we are getting using SVM: Accuracy: 0.82, Precision: 0.86, Recall: 0.69, Specificity: 0.91.
- Values we are getting using KNN: Accuracy: 0.70, Precision: 0.67, Recall: 0.62, Specificity: 0.77.
- Values we are getting using the Decision Tree: Accuracy: 0.77, Precision: 0.73, Recall: 0.73, Specificity: 0.80

VI. CONCLUSION

Table 1. Comparative Analysis of ML Algorithm For GUI-

Process (Method)	Accuracy	Recall /Sensitivity	Specificity	Precision
Logistic Regression	0.80	0.65	0.91	0.85
SVM	0.82	0.69	0.91	0.86
KNN	0.70	0.62	0.77	0.67
Decision Tree	0.77	0.73	0.80	0.73

In this project, we aim to study the prediction of heart disease using machine learning techniques in Python. Given the rising number of deaths caused by heart disease, it is essential to develop systems that can predict heart disease both effectively and accurately. The motivation for this study is to identify the most efficient machine learning algorithm for detecting heart disease. To do this, we compare the accuracy of different algorithms, such as logistic regression and random forest, for predicting heart disease using a dataset from the UCI machine learning repository. The study finds that the Support Vector Machine (SVM) and Gradient Boosting Classifier are the most efficient algorithms, achieving accuracy scores of 82% and 97%

ACKNOWLEDGMENT

The authors are grateful to Dr. Geeta S. Lathkar, Director and Dr. V. G. Saygavi, Vice Principal, MGM's College of Engineering and Technology (MGMCET) and Rev Fr Richard Aloysius Coelho, Director, Father Muller Charitable Institutions (FMCI)for providing continuous encouragement throughout the study. The authors are also thankful to Dr. G. D. Jindal, Professor and Head, Department of Biomedical Engineering, MGMCET for his continuous support and guidance.

REFERENCES

1. A.S. Abdullah and R. R. Rajlaxm, "A data model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Communication . Controls, Apr. 2012, pp. 22–25.
2. Sonam Nekhar, A.M. Karandikar Prediction of Heart Disease Using Machine Learning Algorithms International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication, [Vol-2, Issue-6, June- 2016]. I.S. Jacobs and C.P. Bean, Fine particles, thin films, and exchange anisotropy, in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
3. Kishore 1, Ajay Kumar2, Karan Singh, Maninder Punia4, Yogita Hambir5, Heart Attack Prediction Using Deep Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.
4. A.S. Abdullah and R. R. Rajalaxmi, A data mining appear for predicting coronary heart ailment utilizing self-assertive forest classifier, in Proc. Int. Conf. Afterward Designs Comput. procedures, Communication . Controls, Apr. 2012, pp. 2225.
5. Goldstein BA, Navar AM, Carter RE. Moving backslide regression methods in cardiovascular risk prediction: Applying machine learning to address analytic challenges. Eur. Heart J. 38(23), 1805–1814 (2017).
6. Anuradha P, Dr. Vasantha Kalyani David, "Feature Selection and Prediction of Heart diseases using Gradient Boosting Algorithms", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281- 9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9395819
7. C. Beulah Christian Latha, S. Carolina Sanjeev, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques". Informatics in Medicine Unlocked 16 (2019) 100203 Available online 02 July 2019 2352-9148/ © 2019 Published by Elsevier Ltd. https://doi.org/10.1016/j.imu.2019.100203

© 2024 IJSRET