

Heart Disease Prediction Using Machine Learning

K. Bharathi¹; M. Udaya Naga Bhagyasri²; Garnepudi Maha Lakshmi³; Hanmath Sai Javvaji⁴

¹Assistant Professor, Department of Electronics and Communication Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, Krishna Dist., India.

¹bharathi3663@gmail.com

²Student, Department of Electronics and Communication Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, Krishna Dist., India.

²udayamodugu999@gmail.com

³Student, Department of Electronics and Communication Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, Krishna Dist., India.

³garnepudimahalakshmi284@gmail.com

⁴Student, Department of Electronics and Communication Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, Krishna Dist., India.

⁴hanmathsai@gmail.com

Abstract

CVD is the one of the main reason of death compared to other diseases globally. Nearly 17.9 million people die every year because of this disease. WORLD HEALTH ORGANIZATION predicts that the deaths will increase by 24.5 million in 2030. Heart diseases involve diseases of the guts and blood vessels/artery. Four out of five deaths occur due to heart attacks. Nearly One-third of those deaths occur under the age of 70. The most of the people died in developing countries only. India is also comes under this category. For heart disease identification by examining the symptoms we need vascular specialists, which are very finite number in developing nations. Also, the medical tests/procedures for heart diseases are that's a bit steep; At times out of the range for patient's family. Early recognition is important in these type of diseases with low cost forecasting. In these days ML algorithms are used for forecasting of many other applications. These algorithms can be used for forecasting of CVD. The timely prediction of the heart diseases helped in taking decisions about the patients which resulted reduction of their risks. This project proposes a prediction model to predict whether patient have a heart disease or not by using entered symptoms in webform and to give an awareness on heart disease and some useful tips on heart disease.

Key-words: World Health Organization, Machine Learning, CVD.

1. Introduction

Machine learning disease prediction is a system that predicts diseases based on information provided by users. It predicts the disease of the patient or the user based on the information or the symptoms enter into the web system and gives results based on that information. If the patient isn't much serious and therefore the user just wants to understand the sort of disease, he/she has been through. It's a system which gives the ideas and tips to take care of the health of the user and it provides how to seek out disease using this prediction. so just by entering the symptoms and every one other useful information the user can get to understand the disease he/she is affected by and therefore the health industry also can get enjoy this technique by just asking the symptoms from the user and entering within the system and in only few seconds they will tell the precise and up to some extent the accurate diseases. These kind of ML systems has been implemented by many other organizations, but we intend to make it unique and more useful to users who use this system. This Heart Disease Prediction Using Machine Learning is completely done with the help of Machine Learning algorithms and Python Programming language with Flask Interface for it and also using the dataset that's available previously by the hospitals using that we'll predict the disease. Nowadays's doctors are using many technologies and methodology for not only can identify and diagnose common diseases, but also many deadly diseases.

The exact and accurate analysis is normally attributed to the successful treatment. When doctors fail to make accurate decisions while examining a patient's disease, disease forecasting systems that use ML algorithms can help. The project disease prediction using machine learning is being developed to beat general disease in earlier stages, as we all know in a competitive environment of economic development, mankind has been involved in so much that he/she isn't concerned about. According to research, 40% of people ignore general disease, which leads to harmful disease later. The most common cause of ignorance is a lack of motivation to consult a doctor. In terms of time, people have become so engrossed in their activities that they have no time to convene a meeting or consult a doctor, resulting in a fatal disease. According to research, 70 percent of people in India suffer from general disease, and 25 percent die as a result of early ignorance. The main reason for developing this project is so that a user can sit at their convenience and have a health check-up. The UI is designed in such a way that anyone can easily operate it and have a check-up.

2. Literature Survey

Tom Mitchell states that machine learning as “A computer virus is claimed to learning from the past experience and from some tasks and a few performance on, as measured by, improves with experience”. Machine Learning is combination of correlations and relationships, most machine learning algorithms alive are concerned with finding and/or exploiting interrelationship between datasets. Once Machine Learning Algorithms can identify on certain correlations, the model can either use these relationships to predict future observations or generalize the info to reveal interesting patterns. In Machine Learning there are various types of algorithms such as Regression, Linear Regression, Logistic Regression, Naive Bayes Classifier, Bayes theorem, KNN (K-Nearest Neighbor Classifier), Decision Tress, Entropy, ID3, SVM (Support Vector Machines), K-means Algorithm, Random Forest and etc.

Machine learning made up our minds in 1959 by Arthur Samuel. Machine learning explores the study and building of algorithms which will learn from information and create predictions on information. Machine learning is closely related to procedure statistics, that is in addition focuses on prediction creating through the use of computers. cubic centimeter has sturdy relation with mathematical optimization, that delivers strategies, theory and application domains to the sector. Machine learning is typically combine with data processing, wherever the latter subfield focuses more on exploratory data analysis (EDA) is known as unsupervised learning.

Within this field of data systematic, ML is a sub-method used to difficult models and algorithms that grant themselves to prediction this is known as forecasting analytics. These analytical models is used to allow researchers, data scientists, data engineers, and data analysts to "produce good repeatable choices and outcomes" and expose "hidden patterns" through studying from ancient connection and course in the data.

3. Design Goals

The design goals include a variety of designs that we have implemented in our system to predict disease using machine learning. This system was created using a variety of designs, including a data flow diagram, sequence diagram, class diagram, use case diagram, component diagram, activity diagram, state chart diagram, and deployment diagram. We completed our project after completing these various diagrams and based on these diagrams.

Here are some of the things that this system can do.

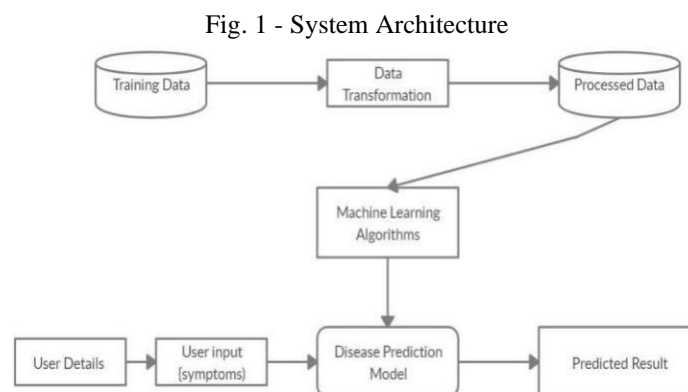
1. Entering Symptoms/Information
2. Disease Prediction

Entering Symptoms: Once user successfully open the system then he/she has to select the symptoms as per the given drop-down menu.

Disease prediction: The predictive model predicts the disease of a person and gives output as might have disease, not disease, depends on the symptoms entered by the user.

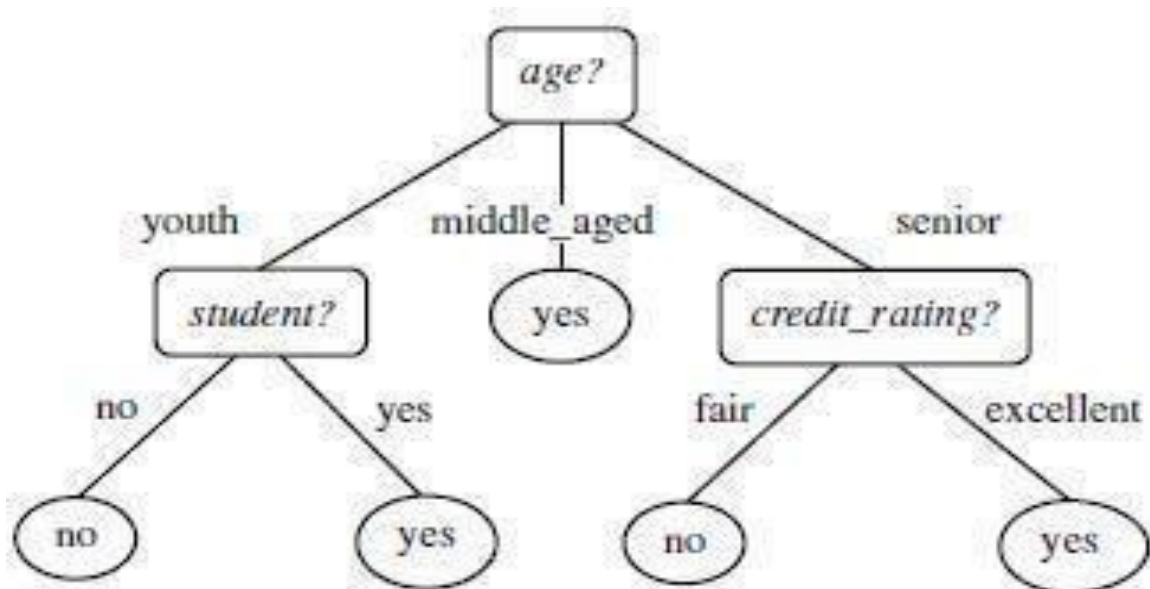
4. System Architecture

Disease forecasting using machine learning predicts the existence of the sickness for the user oriented on various symptoms and the information the user gives a such as glucose level, plasma level and many more such common details through the indications. The planning of the system disease forecasting using machine learning incorporates numerous datasets through that we'll compare the symptoms of the user and predicts it, then the datasets are reworked into the smaller sets and from there it gets classified supported the classification algorithms later on the classified knowledge then processed into the machine learning technologies through that the info gets processed and goes in to the sickness prediction model exploitation all the inputs from the user that's mentioned higher than. Then once user coming information and overall processed information combines and compares within the prediction model of the system and at last predicts the illness. AN design diagram could be a graphical illustration of a group of study ideas, that are a part of AN design, as well as their principles, parts and elements. The diagram explains regarding the system computer code in perception of summary of the system.



5. Implementation

Machine learning disease prediction project aims to overcome common early diseases as we all know in competitive environment of economic development the mankind has involved so much that he/she is not concerned about health according to research, 40% of people ignore common diseases, which then lead to dangerous diseases. The Project “Heart Disease Prediction using Machine Learning” is implemented using python completely. The form is created using html and css with flask integration. After that user needs to enter the details in the form to get respective accurate output. This prediction is basically done with the help of 3 algorithms of machine learning such as Decision Tree, Random Forest, Logistic Regression. When user enter all the symptoms then he needs to press the buttons of respective algorithm, for example there are 3 buttons for 3 algorithms, if user enters all symptoms and presses only Random Forest’s button then the result will be provided only calculating using that algorithm, like this we have used three algorithms to provide more clear picture of the results and user needs to be satisfied with his predicted result on disease.



Decision tree is the supervised learning algorithm and it uses class-labelled training tuples/rows. A decision tree algorithm is a flowchart-like above tree structure and useful for both regression and classification.

Beneath are few presumptions of Decision tree algorithm.

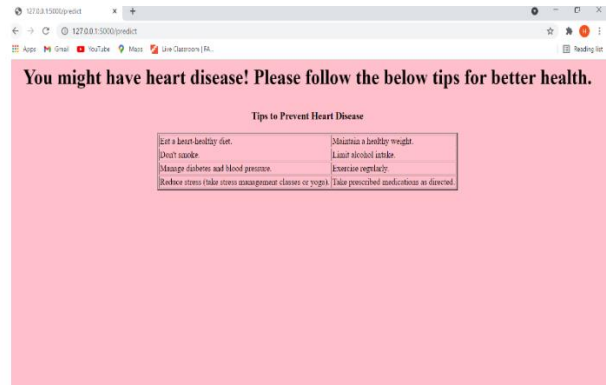
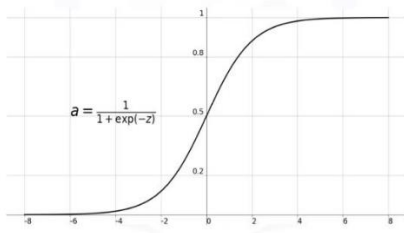
First, the total training set is prepended as the root/node. Attribute values are expected to be categorical. If the values are continuous then they changed to discrete in advance so that the model has been established.

Records and rows are allocated recursively based on attribute/function values. A decision tree employs several algorithms to determine where to divide a node into two or more sub-nodes. The formation of sub-nodes increases/increases the homogeneity of the resulting sub-nodes. In other words, the purity of the node increases in relation to the output variable. The decision tree divides the node into all available variables and then chooses the part with the most homogeneous sub-nodes. For attribute selection, we can use the following criteria: Entropy, Information Gain, Gini Index, Gain Ratio, Reduction in Variance Chi-Square.

The next algorithm we used was Random Forest, which is a supervised learning algorithm that can be used for both regression and classification. When creating decision trees, the training dataset is randomly divided. When splitting the nodes, random subsets of attributes are taken. Because of these two factors, it is referred to as "random." A random forest is the classifier, and it consists of a collection of tree structured classifiers, identically distributed random trees, and each random tree consists of a unit of vote/poll for classification of input. Each tree makes use of the In each tree, In each tree, either entropy or the Gini index is used to split the tree and determine the final label. The final output of each tree is combined and voted on by the weighted values to create the final classifier with the highest accuracy. The random forest algorithm works as follows: A random data set is chosen, which distributes a collection of samples from the training dataset at random while maintaining the class distribution.

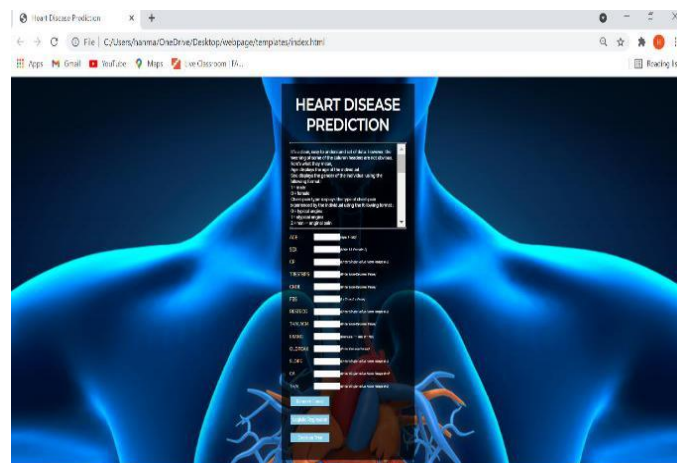
Logistic Regression is a supervised machine learning algorithm that predicts using a logit/sigmoid function. The function returns values ranging from 0 to 1 and logistic regression uses a threshold value based on this value output will be decided either one or zero. If output of sigmoid function is greater than threshold value then our output is one otherwise zero. Loss function in logistic regression is called log loss. The name suggests it's a regression but originally it is useful for the classification problem. Specifically well suited for binary classification.

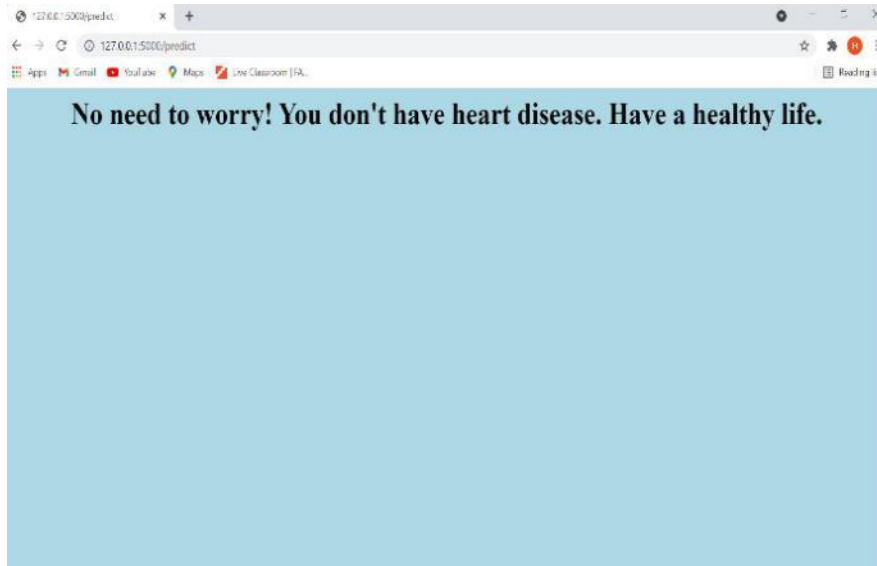
Sigmoid Function



After building ML model with flask integration/interface we can run our application in localhost server. We can also convert this localhost address into public web address using ngrok. Now, We can share this public web link/address to users/friends so that they can use this application in mobile/laptop at everywhere in the world.

6. Result





7. Conclusion and Future Scope

So, Finally I conclude by saying that, this project Disease prediction using machine learning is extremely much useful in everyone's day to day life and it's mainly more important for the healthcare sector, because they're the one that daily uses these systems to predict the diseases of the patients supported their general information and there symptoms that they're been through. Now a day's health industry plays important role in curing the diseases of the patients so this is often also some quite help for the health industry to inform the user and also it's useful for the user just in case he/she doesn't want to travel to the hospital or the other clinics, so just by entering the symptoms and every one other useful information within the form user can get to know the disease he/she is affected by and therefore the health industry also can get enjoy this technique by just asking the symptoms from the user and entering within the system and in only few seconds they will tell the precise and up to some extent the accurate diseases. If health industry adopts this project then the work of the doctors are often reduced and that they can easily predict the disease of the patient. The Disease prediction is to supply prediction for the varied and usually occurring diseases that when unchecked and sometimes ignored can turns into fatal disease and cause lot of problem to the patient. We can update this project in future by adding more attributes to the dataset and more interactive to the users and can also be done as a mobile application. we will modify the system by connecting it to the hospital's database.

References

- Savita ganga sharma (2020).” A Review on machine learning techniques for prediction of cardiovascular diseases”. *2020 Sixth international conference on parallel, Distributed and Grid computing*.
- Singh, A., & Kumar, R. (2020). “Heart Disease Prediction Using Machine Learning Algorithms”. *2020 International Conference on E*.
- “Ali, Liaqat, "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." *IEEE Access* 7(2019): 54007-54014.
- “Santhana Krishnan. J, Geetha S. "Prediction of Heart Disease Using Machine Learning Algorithms", *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*.
- “Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques." *IEEE Access* 7(2019): 81542-81554.
- Singh, Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh, "Heart Disease Prediction System Using Random Forest", *International Conference on Advances in Computing and Data Sciences*. Springer, Singapore, 2018.
- Purushottam, Kanak Saxena and Richa Sharma, "Efficient heart disease prediction system. "*Procedia Computer Science* 85(2016): 962-969.