

SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

EDIA

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

 \bigcirc

e-ISSN: 2319-8753, p-ISSN: 2320-6710 www.ijirset.com | Impact Factor: 8.118



Volume 11, Issue 5, May 2022

DOI:10.15680/IJIRSET.2022.1105027

Cardiovascular Disease Prediction System

Shraddha Deshpande¹, Prajyot Dhumal², Pallavi Kavthale³, Prof. Vijay Sonawane⁴

Student, Dept. of Computer Engineering, JSPM's Bhivarabai Sawant Institute of Technology & Research, Pune,

Maharashtra, India

Assistant Professor, Dept. of Computer Engineering, JSPM's Bhivarabai Sawant Institute of Technology & Research,

Pune, Maharashtra, India

ABSTRACT: The aid domain is one of the distinguished analysis fields within the current situation with the fast improvement of technology and knowledge. it's troublesome to handle the massive quantity of information of the patients. it is easier to handle this knowledge through huge knowledge Analytics. There square measure tons of procedures for the treatment of multiple diseases across the planet. Machine Learning is Associate in Nursing rising approach that helps within the prediction and designation of a unwellness. This paper depicts the prediction of unwellness supported symptoms victimisation m achine learning. Machine Learning algorithms like Naive Thomas Bayes, call Tree and Random Forest square measure utilized on the provided dataset and predict the unwellness. Its implementation is finished through the python programing language. The analysis demonstrates the most effective formula supported their accuracy. The accuracy of Associate in Nursing formula is decided by the performance of the given dataset.

KEYWORDS: knowledge Analytics, Machine Learning, Naïve Thomas Bayes, call Trees, Datasets, Prediction.

l. INTRODUCTION

At present, once one suffers from a specific unwellness, then the person should visit a doctor that is long and expensive too. Also, if the user is out of reach of doctors and hospitals it should be troublesome for the user because the unwellness cannotbe known. So, if the higher than method will be completed victimisation an automatic program which will save time furthermore as cash, it may well be easier forthe patient which might create the method easier. There square measure different Hearts connected unwellness Pre diction Systems victimisation data processing techniques that analyse the chance level of the patient. A heart diseasePredictor could be a web-based application that predicts the centre disease of the user with relevance the symptoms given by the user. the centre unwellness Prediction system has knowledge sets collected from totally different health-related sites. With the assistance of the given symptoms, because the use of the web is growing daily, folks square measure invariably curious to understand totally different new things. folks invariably attempt to talk over with the web if any downside arises. folks have access to the web quite hospitals and doctors. folks don't have a direct choice once they suffer from a specific unwellness. So, this method will be useful to the folks as they need access to the web twenty-four hours.

II. LITERATURE SURVEY

1) Design And Implementing Heart Disease Prediction Using Naïve Bayesian. Author: Anjan Nikhil Repaka, Sai Deepak Ravikanti. Abstract:— Data mining, a great developing technique that revolves around exploring and digging out significant information from a massive collection of data which can be further beneficial in examining and drawing out patterns for making business-related decisions. Talking about the Medical domain, implementation of data mining inthis field can yield in discovering and withdrawing valuable patterns and information which can prove beneficial in performing clinical diagnoses. The research focuses on heart disease diagnosis by considering previous data and information. To achieve this SHDP (Smart Heart Disease Prediction) is built via Navies Bayesian in order to predict risk factors concerning heart disease. The speedy advancement of technology has led to the remarkable rise in mobile health technology that is one of the web applications. The required data is assembled in a standardized form. For predicting the chances of heart disease in a patient, the following attributes are being fetched from the medical profiles, these includeage, BP, cholesterol, sex, blood, sugar etc. The collected attributes act as input for the Navies Bayesian classification for predicting heart disease. The proposed approach includes the following stages: dataset collection, user registration and login (Application-based), classification via Navies

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|

Volume 11, Issue 5, May 2022

| DOI:10.15680/IJIRSET.2022.1105027 |

Bayesian, prediction and secure data transfer by employing AES (Advanced Encryption Standard). Thereafter result is produced. The research elaborates and presents multiple knowledge abstraction techniques by making use of data mining methods that are adopted for heart disease prediction. The output reveals that the established diagnostic system effectively assists in predicting risk factors concerning heart diseases.

2) Application of Machine Learning in Disease Prediction. Author: Rahul Preet Singh Kohli, Shariya Arora. Abstract: The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which can provide data that aids medical experts in the early detection of fatal diseases and therefore, increase the survival rate of patients significantly. In this paper, we apply different classification algorithms, each with its own advantage to three separate databases of disease (Heart, Breast cancer, Diabetes) available in the UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modelling using the p-value test. The results of the study strengthen the idea of the application of machine learning in the early detection of diseases.

III. PROJECT REQUIREMENT

• External Interface Requirement

1. User Interface

Application Based Heart Disease Prediction

2. Hardware Interface RAM:

- RAM: 8 GB
- As we are using Machine Learning Algorithm and Various High-Level Libraries Laptop RAM minimum required is 8 GB.
- Hard Disk: 40 GB
- Data Set of CT Scan images is to be used hence minimum 40 GB of Hard Disk memory is required. Processor: Intel i5 Processor
- PyCharm IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required IDE: PyCharm
- Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that make typing feasible and fast. Coding Language: Python Version 3.5
- Highly specified Programming Language for Machine Learning because of availability of HighPerformance Libraries.
- Operating System: Windows 10
- Latest Operating System that supports all types of installation and development Environment

2. Software Interfaces

Operating System: Windows 10 IDE: Spyder Programming Language: Python

• Non-Functional Requirement

I. Performance Requirement

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work efficiently. The Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement. The application is designed in modules where errors can be detected and easily. This makes it easier to install and update new functionality if required.

II. Safety Requirement

The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

e-ISSN: 2319-8753, p-ISSN: 2320-6710 www.ijirset.com | Impact Factor: 8.118



Volume 11, Issue 5, May 2022

DOI:10.15680/IJIRSET.2022.1105027

III. Software QualityAttributes

Our software has many quality attribute that are given below: Adaptability:

This software is adaptable by all users. Availability: This software is freely available to all users. The availability of the software is easy for everyone.

Maintainability:

After the deployment of the project if any error occurs then it can be easily maintained by the software developer.

Reliability:

The performance of the software is better which will increase the reliability of the Software.

User Friendliness:

Since, the software is a GUI application; the output generated is much user friendly in its behaviour. **Integrity:** Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.

Security:

Users are authenticated using many security phases so reliable security is provided.

Testability:

The software will be tested considering all the aspects.

IV. PROPOSED SYSTEM

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

- 1.) Collection of Dataset
- 2.) Selection of attributes
- 3.) Data Pre-Processing
- 4.) Balancing of Data
- 5.) Disease Prediction

Collection of dataset:

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.



よう 😹 IJIRSET | e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|

Volume 11, Issue 5, May 2022

DOI:10.15680/IJIRSET.2022.1105027

Selection of attributes

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, exang, etc are selected for the prediction. The Correlation matrix is used for attribute selection for this model.



Figure: Correlation matrix

Pre-processing of Data

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.



Balancing of Data

Imbalanced datasets can be balanced in two ways. They are Under Sampling

and Over Sampling

(a) Under Sampling:

In Under Sampling, dataset balance is done by the reduction of the size of the ample class. This process is considered when the amount of data is adequate. (b) Over Sampling:

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| <u>www.ijirset.com</u> | Impact Factor: 8.118|



Volume 11, Issue 5, May 2022

| DOI:10.15680/IJIRSET.2022.1105027 |

In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This process is considered when the amount of data is inadequate.



Prediction of Disease

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression, Ada-boost, Xg-boost are used for classification. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction.



IV. SYSTEM DESIGN

Hearf Disease Predictor A Machine Learning Web Application that predictis chances of having heart Disease or not, Built with Flask and Deployed using Heroku. (Note: This model is 82 67% accurate)		
Age	Your age.	
Sex	select online	~
Chest Pain Type		
Resting Blood Pressure		
Serum Cholesterol	A number in range (94-200) mmng	
Fasting Blood Sugar	A number in range [126-564] mg/dl	
Resting ECG Results	select option	`
Max Heart Rate	select option	~
	A number in range [71-202] bpm	

	Volume 11, Issue 5, May 2022 DOI:10.15680/IJIRSET.2022.1105027
Fasting Blood Sugar	select option
Resting ECG Results	select online
Max Heart Rate	A number in range [71-202] bpm
Exercise-induced Angina	····select option
ST depression	ST depression, typically in [0-6.2]
slope of the peak exercise ST segment	select option V
Number of Major vessels	Typically in [0-4]
Thalassemia	select option v
	Predict
	₽ m

V. CONCLUSION

This project aims to predict heart disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e., predicting heart disease. Average prediction accuracy probability of 55 obtained. All theart Disease Predictor was successfully implemented using the grails framewor

REFERENCES

[1] De Magalhães, João Pedro. "How ageing processes influence cancer." Nature Reviews Cancer 13.5 (2013): 357.

[2] Yixuan Li, Zixuan Chen, Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction, Applied and Computational Mathematics. Vol. 7, No. 4, 2018, pp. 212-216. doi: 10.11648/j.acm.20180704.15

[3] Aloraini, Adel. "Different machine learning algorithms for breast cancer diagnosis" International Journal of Artificial Intelligence & Applications 3.6 (2012): 21.

[4] Gupta, P., and P. S. . " nalysis of Machine earning echniques for Breast Cancer Prediction". International Journal of Engineering and Computer Science, Vol. 7, no. 05, May 2018, pp. 23891-5, http://www.ijecs.in/index.php/ijecs/article/view/4071.

[5] Asri, Hiba, et al. "Using machine learning algorithms for breast cancer risk prediction and diagnosis." Procedia Computer Science 83 (2016): 1064-1069

[6] Rohit Arora and Suman "Comparative Analysis of Classification Algorithms on Different Datasets using WEKA," 2012 International Journal of Computer Applications (0975 - 8887) Volume 54- No.13, September 2012.

[7] Kashish Ara Shakil Shadma Anis Mansaf Alam "Dengue disease prediction using weka data mining tool" ar Xiv preprint arXiv:1502.05167< 2015"

[8] Abhishek Taneja "Heart disease Prediction System Using data Mining Techniques", Oriental Journal of Computer science & Technology, Vol. 6, Issue (4), pp. 457-466, 2013.

[9]J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, pp. 1-5-2016





Impact Factor: 8.118





INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com