

Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively

**Rudra A. Godse¹, Smita S. Gunjal², Karan A. Jagtap³, Neha S. Mahamuni⁴,
Prof. Suchita Wankhade⁵**

Student, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India^{1,2,3,4}

Assistant Professor, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India⁵

Abstract: Using Machine learning, our project proposes disease prediction system. For small issues, the users need to go in person to the hospital for check-up that is longer intense. Also handling the telecom entails appointments is kind of agitated. Such a tangle may be solved by Disease prediction application by giving correct steerage relating to healthy living. Over the past decade, the utilization of the particular disease prediction tools alongside the regarding health has been magnified because of a range of diseases and fewer doctor-patient magnitude relation. Thus, during this system, we have a tendency to area unit concentrating on providing immediate and correct disease prediction to the users concerning the symptoms they enter alongside the severity of disease expected. Best appropriate rule and doctor consultation are given during this project. For prediction of diseases, totally different machine learning algorithms area unit wont to guarantee fast and correct predictions. In one channel, the symptoms entered are crosschecked with the information. Further, can be preserved within the information if the symptom is new that its primary work is and therefore the different channel will offer severity of disease expected. A web/android application is deployed for user for straightforward moveableness, configuring and having the ability to access remotely wherever doctors cannot reach simply. usually users don't seem to be privy to all the treatment relating to the actual disease, this project additionally appearance forward to providing medication and drug consultation of disease expected. Therefore, this arrangement helps in easier health management.

Keywords: Machine Learning, KNN algorithm, SVM, Decision Tree Algorithm, Naïve Bayes Algorithm, Django, Python, etc.

I. INTRODUCTION

The Earth is passing through a violet patch of technology, wherever there's increasing demand of intelligence and accuracy behind it. Today's individuals area unit more likely addicted to web however they're not involved about their personal health. during this twenty first Century humans area unit encircled with technology as they're the constituent of our day to day life cycle. With this we tend to area unit invariably specializing in the health for ourselves and our earned valuables severally. individuals avoid to travel in hospital for tiny downside which can become a significant malady in future. Establishing question answer forums is turning into a straightforward thanks to answer those queries instead of browsing through the list of probably relevant document from the net. Our basic plan is to develop a system which can predict and provides the main points of the disease foreseen together with its severity that as symptoms area unit given as input by the user. The system can compare the symptoms with the datasets provided within the information. If the symptom matches the datasets then it ought to raise alternative relevant symptoms specifying the name of the symptom. If not, the symptom entered ought to be notified as wrong symptom. once this a prompt can come back up asking whether or not you would like to still save the symptom within the information. If you click on affirmative, it'll be saved within the information, if not it'll attend the recycle bin. the most feature are the machine learning, during which we'll be mistreatment algorithms like Naïve Thomas Bayes algorithmic rule, K-Nearest algorithmic rule, Decision Tree algorithmic rule, Random Forest algorithmic rule and Support Vector Machine, which can predict correct disease and additionally, can realize that which algorithmic rule provides a quicker and economical result by comparatively-comparing.

II. LITERATURE SURVEY

[1]This paper contributes the correlative application and analysis of distinct machine learning algorithms in the R software which gives an immediate mechanism for the user to use the machine learning algorithms in R software for forecasting the cardiovascular diseases.

[2]This study aims to understand support vector machine and use it to predict lifestyle diseases that an individual might be susceptible to.

[3] In this paper, two supervised data mining algorithms were applied on the dataset to predict the possibilities of having heart disease of a patient, were analysed with classification model namely Naïve Bayes Classifier and Decision tree classification. The Decision tree model has predicted the heart disease patient with an accuracy level of 91% and Naïve Bayes classifier has predicted heart disease patient with an accuracy level of 87%.

In [4], two different data mining classification techniques were used for the prediction of various diseases and their performance was compared in order to evaluate the best classifier. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for Medical applications.

Heart disease is one of the leading causes of deaths worldwide and the early prediction of heart disease is very important. In this study prove that the proposed new algorithm achieves a highest accuracy compare with another algorithm [5].

In this paper [6], it is presumed that albeit most analysts are utilizing diverse classifier methods, for example, Neural system, SVM, KNN and twofold discretization with Gain Ratio Decision Tree in the conclusion of coronary illness, applying Naïve Bayes and Decision tree with data pick up counts gives better outcomes in the finding of coronary illness and better exactness when contrasted with different classifiers.

III. EXISTING SYSTEM

In the existing system the info set is often little, for patients and diseases with specific conditions. These systems are principally designed for the additional prodigious diseases like cardiovascular disease, Cancer etc. The pre-selected characteristics could generally not satisfy the changes within the malady and its influencing factors that may lead to quality in results. As we tend to sleep in ceaselessly evolving world, the symptoms of diseases conjointly evolve over a course of your time. conjointly most of the present systems build the users await long periods by creating them answer prolonged questionnaires.

IV. PROPOSED SYSTEM

This paper proposing such a system which will flaunt a simple and elegant User Interface and also be time efficient. In order to make it less time consuming we are aiming at a more specific questionnaire which will be followed by the system. Our aim with this system is to be the connecting bridge between doctors and patients. The main feature will be the machine learning, in which we will be using algorithms such as Naïve Bayes Algorithm, K-Nearest Algorithm, Decision Tree Algorithm, Random Forest Algorithm and Support Vector Machine, which will help us in getting accurate predictions and Also, will find which algorithm gives a faster and efficient result by comparatively-comparing. Another feature that our system will comprise of is Doctor's Consultation. After delivering the results, our system will also suggest the user to get a doctors consultation on this report.

Advantages of Proposed System :

- I. Simple UI
- II. Time Efficient
- III. Doctor Consultation
- IV. Cost Efficient
- V. Specific Questionnaire

V. SYSTEM ARCHITECTURE

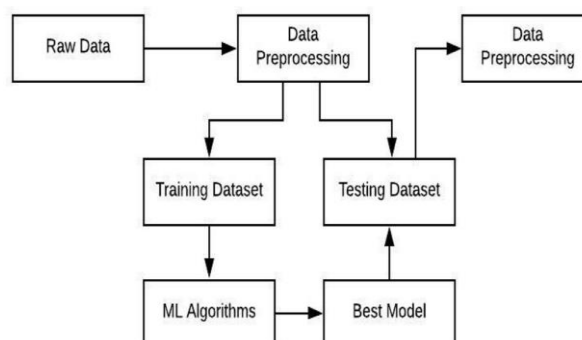


Fig. 1. ML System Architecture

As shown in Fig. 1., the information from the first dataset is passed onto the primary phase I.e. information pre-processing. In information pre-processing this information is then clean of all redundancies, missing values etc. The new clean information is acceptable coaching completely different algorithmic models thereon.

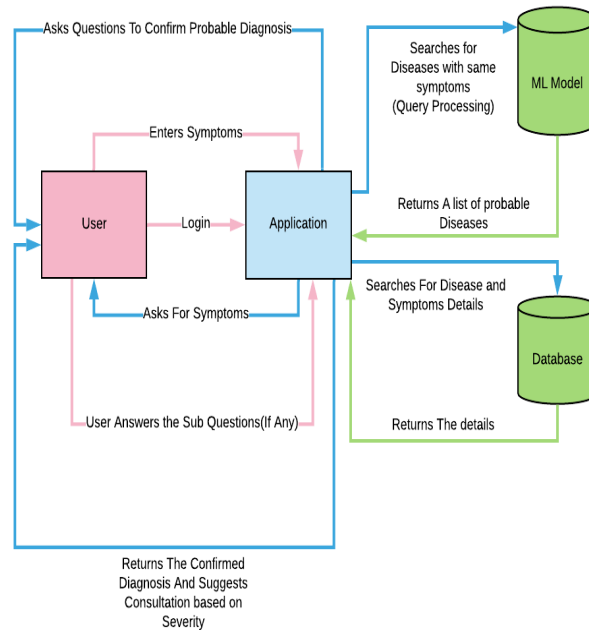


Fig. 2. Data Flow Diagram of Working

VI. MATHEMATICAL MODEL

Input x, Output y
Therefore, $y = f(x)$

I: Set of Inputs = {I1, I2, I3}

Where,

I1: Login

I2: User Details

I3: Symptoms

Output = {O1, O2, O3}

Where,

O1 = The Predicted Disease.

O2 = Doctors Consultation.

O3 = The Disease Description.

Function = {F1, F2, F3, F4}

Where,

F1 = The user registers into the application.

F2= The application asks the user to enter the personal details with medical background.

F3= The user logs into the application.

F4= Enters the symptoms.

F5= Disease is predicted according to the symptoms.

Let 'S' be the solution perspective of the class

$S = \{s, e, i, o, f, \text{success, failure}\}$

where,

s = Start of program.

e = The end of program.

i = Personal medical information of the patient and symptoms.

o = Disease detected.

success = Accurate disease detected.

failure = Incorrect disease predicted.

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

VII. METHOD DESCRIPTION

In this paper, We are using total five algorithms for predicting the disease. Which are Naïve Bayes, SVM, KNN, Decision Tree and Random Forest Algorithm.

NAÏVE BAYES.

Naive mathematician classifiers are a set of classification algorithms supported Bayes' Theorem. it's not one formula however a family of algorithms wherever all of them share a standard principle, i.e. each try of options being classified is freelance of every different.

SVM.

Support Vector Machine or SVM is one among the foremost standard supervised Learning algorithms, that is employed for Classification further as Regression issues. However, primarily, it's used for Classification issues in Machine Learning. The goal of the SVM rule is to make the simplest line or call boundary that may segregate n-dimensional house into categories in order that we will simply place the new datum within the correct class within the future. This best call boundary is termed a hyperplane. SVM chooses the acute points/vectors that facilitate in making the hyperplane. These extreme cases square measure known as support vectors, and therefore rule is termed as Support Vector Machine.

KNN.

K-nearest neighbors (KNN) rule could be a variety of supervised metric capacity unit rule which may be used for each classification yet as regression prophetic issues. However, it's primarily used for classification prophetic issues in trade. the subsequent 2 properties would outline KNN well –

- Lazy learning algorithm – KNN could be a lazy learning rule as a result of it doesn't have a specialised coaching section and uses all the information for coaching whereas classification.
- Non-parametric learning algorithm – KNN is additionally a non-parametric learning rule as a result of it doesn't assume something regarding the underlying knowledge.

Decision Tree Algorithm.

In medicative selection like arrangement, diagnosis there are varied circumstances wherever selection should be created with success and reliably. cheap easy basic leadership models with the probability of programmed learning ar the foremost fitting for playacting such undertakings. selection trees are a solid and successful basic leadership procedure that furnish high grouping accuracy with an easy portrayal of accumulated learning and that they are utilised as an area of varied zones of restorative basic leadership.

Random Forest Algorithm.

Random Forest could be a learning methodology that operates by constructing multiple call trees. the ultimate call is created supported the bulk of the trees and is chosen by the random forest. There square measure heaps of advantages to victimisation Random Forest algorithmic rule, however one amongst the most benefits is that it reduces the chance of overfitting and therefore the needed coaching time. in addition, it offers a high level of accuracy. Random Forest algorithmic rule runs with efficiency in massive information bases and produces extremely correct predictions by estimating missing data.

VIII. CONCLUSION

This paper offers analysis of multiple researches exhausted this field. Our planned System aims at bridging gap between Doctors and Patients which can facilitate each categories of users in achieving their goals. this technique provides support

for multiple sickness prediction mistreatment completely different Machine Learning algorithms. this approach of the many systems focuses solely on automating this method that lacks in building the user's trust within the system. By providing Doctor's recommendation in our system, we have a tendency to guarantee user's trust aspect by aspect making certain that the Doctor's won't feel that their Business is obtaining affected because of this technique.

REFERENCES

- [1]. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms" Dinesh Kumar G., Santhosh Kumar D.(2018).
- [2]. "A Proposed Model for Lifestyle Disease Predict Vectorion Using Support Machine" Mrunmayi Patil, Vivian Brian Lobo, Pranav Puranik, Aditi Pawaskar, Adarsh Pai, Rupesh Mishra U.G. Student, Assistant Professor (2018).
- [3]. 3. "Prediction of Heart Disease using Machine Learning Algorithm" Mr. Santhana Krishnan J, Dr. Geetha. S. (2018).
- [4]. "Multi Disease Prediction using Data Mining Techniques" K. Gomathi, Dr. D. Shanmuga Priyaa (2017).
- [5]. "Analysis of Heart Disease Prediction Using Datamining Techniques" S.Sharmila, Dr. M. P. Indra Gandhi (2017).
- [6]. "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" G.Parthiban, S.K.Srivasta, International Journal of Applied Information Systems (IJ AIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3– No.7, August 2012.
- [7]. Jaymin Patel, Prof.Tejal Upadhyay, Dr. Samir Patel "Heart disease prediction using Machine learning and Data Mining Technique" Volume 7.Number1 Sept 2015March 2016.
- [8]. G. Parthiban, S. K. Srivasta "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" International Journal of Applied Information Systems (IJ AIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3– No.7, August 2012.
- [9]. Thenmozhi. K and Deepika. P, Heart Disease Prediction using classification with different decision tree techniques. International Journal of Engineering Research & General Science, Vol 2(6), pp 6-11, Oct 2014.
- [10]. "Disease Prediction Using Machine Learning Over Big Data" Vinitha S, Sweetlin S, Vinusha H and Sajini S (2018).
- [11]. "Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning" Reddy Prasad, Pidaparathi Anjali, S. Adil, N. Deepa (2019)
- [12]. "Heart Disease Prediction Using Data Mining Techniques" H. Benjamin Fredrick David and S. Antony Belcy(2018).
- [13]. "Multi Disease Prediction Using Data Mining Techniques" K. Gomathi , Dr. D. Shanmuga Priyaa (2017).