International Journal of Science and Healthcare Research
DOI: https://doi.org/10.52403/ijshr.20211004
Vol.6; Issue: 4; Oct.-Dec. 2021
Website: ijshr.com
Original Research Article *ISSN: 2455-7587*

Disease Prediction using Naïve Bayes - Machine Learning Algorithm

Chandrasekhar Rao Jetti[1], Rehamatulla Shaik[2], Sadhik Shaik[3], Sowmya Sanagapalli[4]

[1] Associate Professor, [2,3,4] UG Students,
Dept. of ECE, Bapatla Engineering College, Bapatla, A.P., India

Corresponding Author: Chandrasekhar Rao Jetti

ABSTRACT

It can occur on many occasions that you or a loved one requires urgent medical assistance, but they are unavailable due to unforeseen circumstances, or that we are unable to locate the appropriate doctor for the care. As a result, we will try to incorporate an online intelligent Smart Healthcare System in this project to solve this issue. It's a web-based programmed that allows patients to get immediate advice about their health problems.

The aim of the smart healthcare system is to create a web application that can take a user's symptoms and predict diseases, as well as serve as an online consultant for various diseases. We created an expert system called Smart Health Care System, which is used to make doctors' jobs easier. A machine examines a patient at a basic level and recommends diseases that may be present. It begins by inquiring about the patient's symptoms; if the device is able to determine the relevant condition, it then recommends a doctor in the patient's immediate vicinity. The system will show the result based on the available accumulated data. We're going to use some clever data mining techniques here. We use several intelligent data mining techniques to guess the most accurate illness that could be associated with a patient's symptoms, and we use an algorithm (Naive Bayes) to map the symptoms with potential diseases based on a database of many patients' medical records. This system not only makes doctors' jobs easier, but it also benefits patients by getting them the care they need as soon as possible.

Keywords: Disease Prediction, Naïve Bayes, Machine Learning Algorithm, Smart Healthcare System

1. INTRODUCTION

When anyone is already infected with an illness, they must see a doctor, which is both time consuming and expensive. It can also be difficult for the user if he or she is unable to contact a doctor or a hospital because the illness cannot be detected. So, if the above procedure can be done using automated software that saves time and money, it might make the process simpler for the patient. Other Heart Disease Prediction Systems use data mining techniques to analyses the patient's risk level.

A smart healthcare system is a web-based programmed that predicts a user's illness based on the symptoms they have. Data sets from various health-related websites have been compiled for the smart healthcare system. The consumer would be able to determine the likelihood of a disease based on the symptoms provided by this method.

People are always curious to learn new things, particularly as the use of the internet grows every day. When an issue occurs, people often want to look it up on the internet. Hospitals and physicians have less access to the internet than the general public. When people are afflicted with an illness, they do not have many options. As a result, this device can be beneficial.

The smart healthcare system is a project that provides end-user assistance and online consultation. In this paper, we suggest a framework that allows users to get online health advice from an intelligent health care system. Various symptoms and diseases were fed into the system. Users may share their symptoms and problems with the system. It then analyses the user's symptoms to see if there are any diseases that may be linked to it.

In this paper, we use intelligent data mining techniques to guess the most reliable suspected disease that could be linked to the patient's symptoms, and we use the algorithm (Naive Bayes) to map the symptoms to possible diseases. This method not only makes doctors' jobs easier, but it also benefits patients by getting them the care they need at the earliest possible stage.

The paper is organized as follows; Section 2 highlights the basic objective, its purpose and the algorithm used to develop this system. Section 3 reveals minimum hardware and software requirements and its usage inside the project. Section 4 gives the proposed design relation between each class using ER Diagrams and Conclusions are discussed in Section 5.

2. Brief overview of Smart Health Care

2.1 Objectives

The key goals of a smart healthcare system are twofold.

a. Create a Nave Bayes Classifier that classifies the disease based on the user's feedback.
b. To build a web interface framework for disease prediction.

2.2 Purpose and Scope

2.2.1 Purpose

There are several methods for disease prediction. However, heart-related diseases have been studied, and a risk level has been calculated. However, such methods are not commonly used for disease prediction in general. As a result, the smart healthcare system aids in the prediction of general diseases. In certain cases, you or a family member may need urgent medical assistance, but physicians are unavailable due to unforeseen circumstances, or we may be unable to locate the appropriate doctor for the care. To address this problem, we will attempt to incorporate an online intelligent Smart Healthcare System in this project. It's a web-based programme that allows patients to get immediate advice about their health problems.

2.2.2 Scope

The aim of this project is to develop a web platform that can predict disease occurrences based on various symptoms. The consumer will search for diseases based on their probabilistic figures by selecting different symptoms.

2.3 Limitations

The Smart Healthcare System has the following limitations:

a. All information on this project is intended for personal educational and medical guidance only; patients should contact their physicians for accurate diagnosis and treatment.
b. The patient's medical history has not been taken into account.
c. Doesn't allow you to create data mining dimensions.

2.4 Algorithm Used: Naive Bayes Algorithm

The Naive bayes algorithm is a classification algorithm that uses Bayesian techniques and is based on the Bayes theorem in predictive modelling. [1] Some algorithms are more computationally intensive than this one. As a result, it can be used to quickly generate mining models to find relationships between input columns and predictable columns.

The Naive Bayes algorithm is primarily used in the creation of classifiers. [2]

The categorical class labels are predicted using certain classifiers. Classifiers, on the other hand, are used to determine which class the given inputs belong to.

For the advancement of the expectation framework, the proposed smart healthcare system framework employs a data mining technique known as "Naive Bayes classifier." [3] This system contains a greater number of data indexes and characteristics that are genuinely collected from expert data in order to determine the exact symptom expectation. Some AI and data mining techniques are based on the "Naive Bayes or Bayes" Rule. The norm is used to create models of precognitive abilities. It benefits from the "proof" by determining the relationship between the goal (i.e., subordinate) and other variables.

The Bayes theorem and Bayes approximation are used in the Naive Bayes algorithm. Bayes theorem P(c|x) = P(x|c) P(c)/P(x), where, P(c|x) is the posterior probability of class (target) given predictor. P(c) is the prior probability of class. P(x|c) is the probability of predictor given class likelihood. P(x) is the prior probability of predictor.

Naive Bayes algorithm is a simple technique which is used for developing the models that are used to assigns class labels to problem instances. The class labels are drawn from finite set. Naïve Byes algorithm is not a single algorithm, but it is a family of algorithm based on a common principle.

This principle states that the value of each feature is independent of values of other features of all Naive Bayes classifiers. There are many probability models, out of which the naïve byes algorithm is efficiently trainee The Naive Bayes is a straightforward method for creating templates for assigning class labels to optimization problems. The class labels are chosen from a finite set of options. The Nave Bayes method is a family of algorithms based on a general concept, rather than a single algorithm. [4] The value of each feature of all naive bayes classifiers is proportional to the value of the other features, according to this theory. There are a variety of probability models, but the nave byes algorithm is one of the most effective in supervised training.

2.5 Proposed System

You can find yourself in a situation where you need urgent medical assistance but the doctor is unavailable due to a variety of factors. We created a smart health prediction framework to address the shortcomings of current systems. This method helps users to have their symptoms analyzed in order to predict the condition they are suffering from.

We created an expert system called Smart Healthcare System, which is used to make doctors' jobs easier. A machine examines a patient at a basic level and recommends diseases that may be present. It begins by inquiring about the patient's symptoms; if the device is able to determine the relevant condition, it then recommends a doctor in the patient's immediate vicinity. If the system is unsure, it will show the result based on the available accumulated information.

We use several analytical data mining techniques to estimate the most accurate illness which could be associated only with patient's condition, and an algorithm (Naive Bayes) is applied to map the symptoms with potential diseases based on a database of multiple disease symptoms records. This system not only makes doctors' jobs easier, but it also benefits patients by getting them the care they need as soon as possible.

3. SYSTEM ANALYSIS

3.1 Hardware Minimum Requirements

Development time requirements:

- Processor : Intel I5 processor
- Hard Disk : 400 GB
- RAM : 4 GB

Run time requirements:

- Processor : Intel I3 processor
- Hard Disk : 200 GB
- RAM :2 GB

3.2 Software Requirements

- Operating System : Windows 7 and above

- Front end : Html, CSS, java Script
- Back end : Php
- Database : MySQL
- Server : Xampp Server.

3.3 Justification of Platform

- Html is used to create effective web pages, CSS is used to create effective presentations, and java script is a client-side language used for dynamic actions. [5]
- PHP (hypertext pre-processor) is an open source and freely accessible server-side scripting language that is primarily used to create static and interactive websites and web applications. [6]
- MySQL is a free open-source relational database management system that allows us to store large amounts of data in a safe manner while still allowing us to retrieve data quickly. [7]
- XAMPP is a portable and simple solution that can run on a variety of platforms (windows, Linux and Mac). Xampp stands for cross platform, Apache, MySQL, PHP, and Perl, and it enables you to build web applications on your computer's local web server. [8]

4. E-R Diagrams

The ER, or Entity Relational Model, is a computational data model diagram at a high level. The Entity-Relation model is focused on the concept of real-world entities and their interactions. ER modelling aids in the systematic analysis of data specifications in order to create a well-designed database. As a result, completing ER modelling before implementing the database is considered best practice. The relationships of an entity collection stored in a database are depicted in an entity relationship diagram. To put it another way, ER diagrams assist you in explaining the logical structure of databases. An ER diagram appears to be very similar to a flowchart at first glance. The following individuals are part of the smart healthcare system: Patient, Doctor, Admin, Disease, Feedback, Appointment, Notification, and Medicine

The connections are as follows:

A patient search for a disease. A patient's search for a physician. The patient provides suggestions. The patient should schedule an appointment with the doctor. The administrator has the ability to introduce new diseases. Doctors will prescribe the latest medications. Doctors have access to the data provided by the system. Doctors will view their schedules. The input can be checked by the administrator.

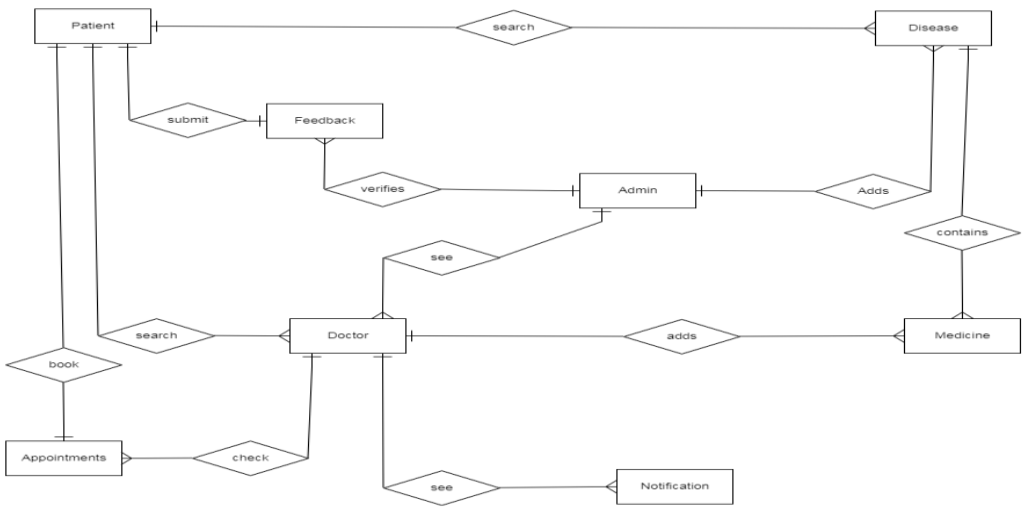

Fig. 1: Entity Relationship diagram for smart healthcare system**.**

4.1 Data Flow Diagrams / UML

A data-flow diagram (DFD) is a visual representation of a process or system's data flow. There are no control flows, decision rules, or loops in a data-flow diagram. The functions, or processes, that collect, manipulate, store, and distribute data between a system are graphically represented by the DFD. The level 0 data flow diagram, also known as a background diagram, is the simplest view of the entire system.

All users will be able to appreciate the level o data flow diagram.

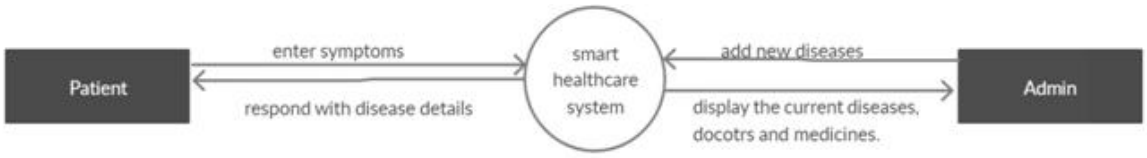

Fig. 2. Level 0 data flow for smart healthcare system.

A level 1 data flow diagram (DFD) is more comprehensive than a level 0 DFD, but not quite as detailed as a level 2 DFD. The key processes are broken down into subprocesses.

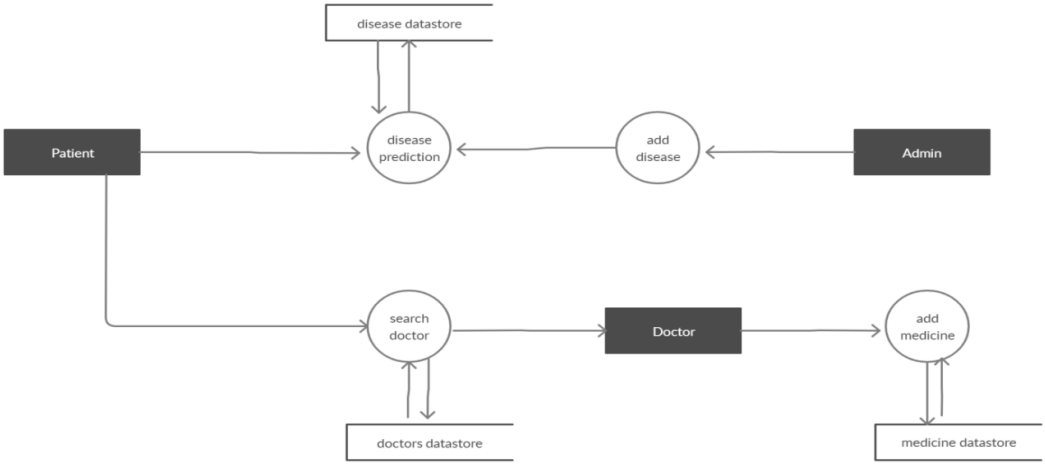

Fig. 3. Level 1 data flow for smart healthcare system.

5. CONCLUSION

The core idea behind the project is to propose a system that allows users to get instant guidance on their health issues. The application takes the input of various symptoms from the patient, does the analysis of the entered symptoms, and gives appropriate disease prediction.

Our system allows the users to get analysis on the symptoms they give for predicting the disease they are suffering from. We have designed a new health prediction system, which is an online system, and various patients from any locations can view it.

Data mining can be beneficial in the field of medical domain. However, privacy, security and unable to log into the account are the big problems if they are not addressed and resolved properly, which can be solved using Blockchain Technology.

Acknowledgement: None

Conflict of Interest: None

Source of Funding: None

REFERENCES

1. Irshad Ullah, “Data Mining Algorithms And Medical Sciences”,IJCSIT, Vol 2, No 6, December 2010
2. Amit Kumar Das, Aman Kedia,“Data mining techniques in Indian Healthcare: A Short Review”, 2015 International Conference on Man and Machine Interfacing (MAMI), 978-1-5090-0225-2/15
3. T. Devasia, T. P. Vinushree and V. Hegde, "Prediction of students performance using

Educational Data Mining", *Proceedings of IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 50-100, 2016.

4. W. Song, C. H. Li and S. C. Park, "Expert Systems with Applications Genetic algorithm for text clustering using ontology and evaluating the validity of various semantic similarity measures", *Expert Syst. Appl*, vol. 36, no. 5, pp. 9090-9110, 2009.
5. Seung-Jin Lim, Yiu-Kai Ng and Xiaochun Yang, "Integrating HTML tables using semantic hierarchies and meta-data sets", *International Database Engineering and Application Symposium (IDEAS)*, 2002.
6. Zurkiewicz, M Milosz, "Selecting a PHP Framework for a web application project-the method and case study", 2013.
7. R. Lawrence, "Integration and Virtualization of Relational SQL and NoSQL Systems Including MySQL and MongoDB", *Proceedings of International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 250-300, Mar. 2014.
8. Sylviane Levi, "Quality Requirements for Multimedia Interactive", *Journal of Software Engineering and Applications Sci Res*, 2013.

How to cite this article: Jetti CR, Shaik R, Shaik S et.al. Disease prediction using Naïve Bayes - Machine learning algorithm. *International Journal of Science & Healthcare Research.* 2021; 6(4): 17-22. DOI: *https://doi.org/10.52403/ijshr.20211004*

******