

# Health Prediction and Medical Diagnosis using Naive Bayes

Himdeep Bohra<sup>1</sup>, Amol Arora<sup>2</sup>, Piyush Gaikwad<sup>3</sup>, Rushabh Bhand<sup>4</sup>, Manisha R Patil<sup>5</sup>

Student, Computer Department, Smt. Kashibai Navale College of Engineering, Pune, India<sup>1,2,3,4</sup>

Assistant Professor, Computer Department, Smt. Kashibai Navale College of Engineering, Pune, India<sup>5</sup>

**Abstract:** In today's modern world, many of us are using computer and its applications for various works. Lot of people are using it in their daily life. Use of computer applications in medical field is also significant. It can provide user an interface for medical guidance. Platform independent system can help people to use it on their computers and will provide users an instant guidance on their health issues. System will use Naïve Bayes algorithm and depending on the symptoms will predict the diseases and for normal person it will predict the daily hygiene diet and routines which he can follow. Users will also be able to contact the specialist doctors nearby. It will help users for easy medical treatment and diagnosis.

**Keywords:** Naïve Bayes, NetBeans, Java Swing, Data mining, classification, standalone application.

## I. INTRODUCTION

In modern world people are using computing resources in many forms like laptops and personal computers. Many people on daily basis are suffering from various diseases and requires respective treatment. Due to modern life people are not paying proper attention to their health both mental health and physical health. Burgeoning of computer resources helps people in their various works. Many people spend time online for searching health information.

A platform independent system can help users to get respective assistance for their health related issues. It may not need online assistance always. In a survey [1] it has been found out that average U.S. consumer spends around 52 hours online searching for the wellness knowledge but visits the doctor only 3 times per year in 2013.

Users want to get the health treatment easily in the comfortable way and wants to get the information of the doctors as easily as possible. An application should provide the doctor and patient interaction in the easiest way. Platform independent application can provide the comfort to the users of using them on their respective machines and an interactive Graphical User Interface (GUI) will indeed help in achieving it.

Health prediction system will be having the two cases, one is for the use of the sick patient and other one is for the healthy person. For sick person application will load the other module and respectively interact with the user. For sick person the system will be asking the symptoms and respectively predict the diseases he or she may be suffering from. Module for the healthy person will be showing the respective healthy routines and diet he or she can follow on daily basis. Person looking for better fitness and stamina will have to provide input in terms of height, weight, age and their respective medical history.

Thus both type of users are able to get the necessary medical help and both of them are able to contact the doctors nearby their area. Users with online help can also be able to interact with specialist doctors and thus can provide quick resolution the problems.

Many official portals are run by sectors, organizations and professional health providers. One more is community based services [2] which provide interactive platforms where health seekers can ask questions and doctors provide their respective trustworthy answers. However these services require constant internet support and are usable for urban sector people only. Secondly these system are time consuming in for health seekers to get their questions resolved and lastly doctors have to take care of lot of work load. Application which are platform independent can resolve many of these issues and can also help in automatic prediction of their questions. Naïve Bayes classifier is a family of classifiers which are probabilistic in nature and based on applying Bayes theorem with strong independence assumptions between the parameters. Bayes theorem can be used for prediction of the diseases as a classifiers decision rule. The probability of the disease according to its number of symptoms can be predicted using Bayes algorithm.

While the standalone application will provide the user an interaction for taking the required inputs, that is symptoms. Bayes classifier considers each of the parameters to contribute independently to the probability and does not consider correlation between parameters. Users therefore are able to get their respective medical diagnosis and get respective medication.

## II. USER INTERACTION WITH STANDALONE APPLICATION

When a user is interacting with an online portal or some interactive platform, he has to be online for getting answers to his queries. While a platform independent application can help people to install it one time on their personal computer or laptops to use it efficiently. Application only needs to be connected online for some urgent works rarely which can be done easily. These standalone applications are application which does not depend on the operating system and its processes. It creates its own process and provides an interface to the user.

Java Swing applications provide Graphical User Interface(GUI) for the application to execute and take queries from user. Java Swing is an Application programming interface for providing a graphical user interface for java programs or java code. Java can be used to build the platform independent application and for this any Integrated Development Environment (IDE) can be used like NetBeans. These IDE can provide variety of tools and environment to make such applications and these applications can also be connected to internet by providing respective methods.

User can be a patient or a normal person who is looking for the health related enquiry. As shown in Fig1. User has to only login to the application and then provide his symptoms as input and then by doing the keyword searching respective symptoms are classified and then Bayes theorem is applied for the matching of the disease as per the symptoms and the maximum related disease will be come as output. Diseases and their respective symptoms are stored in the database and it has to be fetched according to algorithm result.

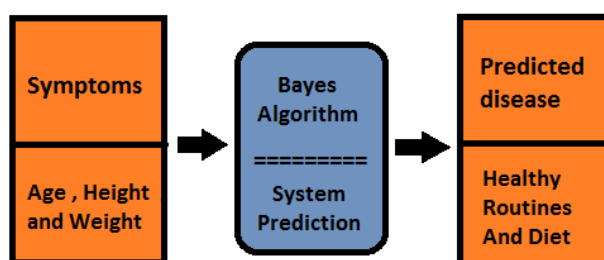


Fig.1 Application Conceptual Diagram

Every working computing model needs data to process and produce the output. Data Mining is the computational process of discovering some relation between data from data sets or some patterns from the data sets. In application processing user gives input and the classification is done for the symptoms and then using the Bayes theorem it will predict the output for the user. While classification is done the symptoms are divided into various classes and respective class will then show the resultant disease. Application model will also ensure user's privacy and data security. It will provide the network security while connecting online and having interaction with doctor and it will also provide security to user while using it on their devices like they have to login and provide correct credentials.

## III. CLASSIFICATION AND DISEASES PREDICTION USING BAYES ALGORITHM

In data mining of [3] of general health records provides sensitive and important data on relations of symptoms and disease. Health examination records can provide data as it has not been explored fully and only few studies on risk prediction of chronic diseases warning system has been done [3]. Person will provide his or her symptoms directly to the system and thus from bayes theorem it will be able to predict the disease. In earlier approach they have used classification which have not clearly shown the diseases that are affecting.

Unlabelled data classification generally uses the semi-supervised learning that is from both labelled data and unlabelled data [3]. While classification is done on labelled data helps in training the system consistently. Training labelled dataset will be used to know whether the system is providing correct prediction or not. Multi-class learning model [3] uses activity recognition and method trains binary probabilistic base classifiers each trained on the positive datasets and this helps making the class decision. Class having the good probabilistic value will be given in the output as a result.

In our system user will provide symptoms when he is sick and age, weight and age in normal and then then the system will be using the Bayes algorithm for the prediction of the disease based on the probabilistic values and it will use the dataset training.

### A. Processing Input and respective classification

The input provided will be in query based form and then it will be tokenised properly using the important keywords related to already store symptoms in the database. The first step is [4] tokenization which will be used here as input query from users and it will be converted into tokens. Then the symptoms will be classified basis on the classes which are already in the system, each class is having the symptoms and shows final respective disease. Using the labelled data set, system will be trained and the resultant classification rules can be used.



### B. Bayes Algorithm for disease prediction

In general, machine learning algorithms have to be trained and for this they must be trained on labelled data sets, this is supervised learning and on basis on this, system can provide correct classification and prediction for the symptoms.

In Bayes algorithm the prior probability will be calculated on the basis of the training data set. Prior probability is the probability of disease that may occur on the basis of its symptoms. It will be conducted on the data set.

$$P(X=\text{CHICKENPOX}) = (\text{fatigue, fever \& appetite loss}) / (\text{Overall Symptoms})$$

Above is the conditional probability and it will be used as prior probability in the future prediction. Bayes algorithm will predict the disease basis on the maximum value among the classes. Then that class will represent its disease or will be selected.

Suppose User A want to find whether he is suffering from malaria given that Prior probability of it is as

$$P(X=\text{Malaria}) = (\text{fatigue, fever, shivering, sweating, muscle pain}) / (\text{overall symptoms})$$

Now a person is having fever and muscle pain then,

$P(Y=\text{fever and muscle pain})$ , then given that the probability of having Malaria (from training data set), then what are the chances that this user A is suffering from Malaria, therefore

$$P(Y | X) = P(X | Y) * P(Y) / P(X).$$

$P(X)$  should not be 0 (zero).

$$\text{Or } P(Y | X) = P(Y \cap X) / P(X).$$

$P(X)$  not equal to 0 (zero).

### C. System implementation and interaction

As the user provides symptoms to the application, application first will do the pre-processing tasks such as tokenization of words and selecting only the key tokens from it [4]. Removing of unnecessary words and punctuations will also assist in the easy analysis of tokens. Firstly the user will have to login into the system by providing the correct credentials and will have to provide the medical history of himself. It will help in the correct authorisation of the user and will also lead in the correct diagnosis.

After secure access, user will get the result as the diseases he may be suffering from and will also get the corresponding medication which system will provide according to the one to one relationship established in the database. It will be able to store the updated data by administrator from time to time accordingly. Indexing on the database will provide faster access to the user on his or her computer for the medical diagnosis from time to time. Administrator will maintain the system and will keep it up to date. Doctors can look into the users request and queries and provide the answers. Both of them can also communicate with each other through the online query post and answering. Doctors can login on their module and discuss among themselves for particular disease and its treatment as shown the Fig. 2. Users query after classification is processed through the Bayes algorithm and will output the result as inferred disease and its respective medication. It will the display back the result and will update user's history as shown in Fig. 2.

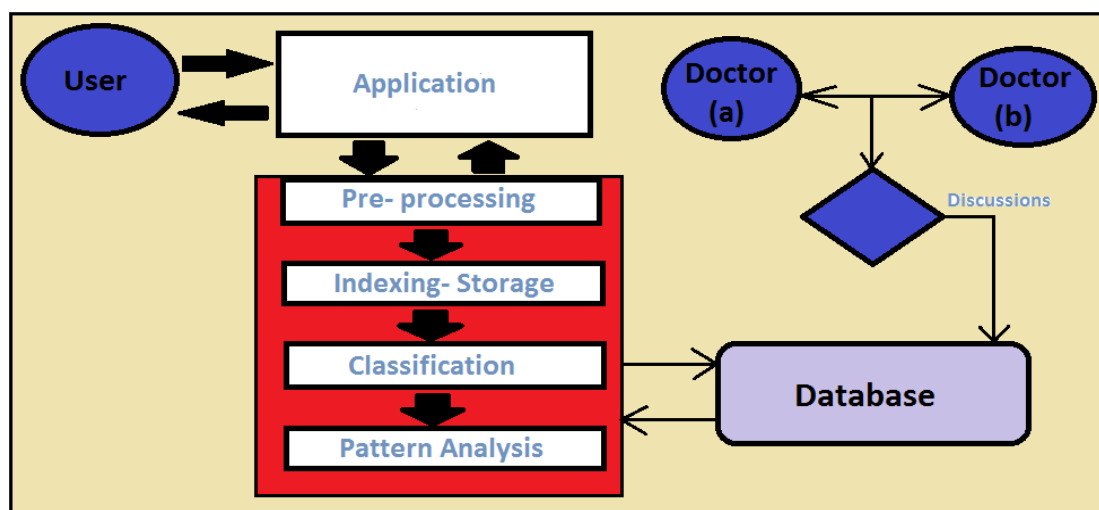


Fig. 2 System implementation diagram

## IV. OUR PROPOSED APPROACH

The approach is utilizing the features of java swing applications and data mining, mainly using the Bayes algorithm and predicting the disease according to the users query input in the form of symptoms. As shown in Fig. 2 the aim of our approach is to deliver a platform independent system to user that would help them in getting medical care easily and



efficiently. To further define the goal of this approach the intended application will provide the user an easy, secure and trustworthy platform for instant medical diagnosis, better advice on healthy routines and diet and better interaction with the specialist doctors. Few features include the online questioning interaction with specialist doctors and calculation of Body Mass Index (BMI) and respectively providing the healthy guidance to stay fit and healthy. This will improve the quality of medical care in both rural and urban area.

The application is platform independent java swing application and will be using database at backhand. Database can be MySQL database which is relational database management system and provides faster access to large amount of data. Once the application is developed it can installed on any platform of computing device. It uses the swing interface which is user friendly and easy to use.

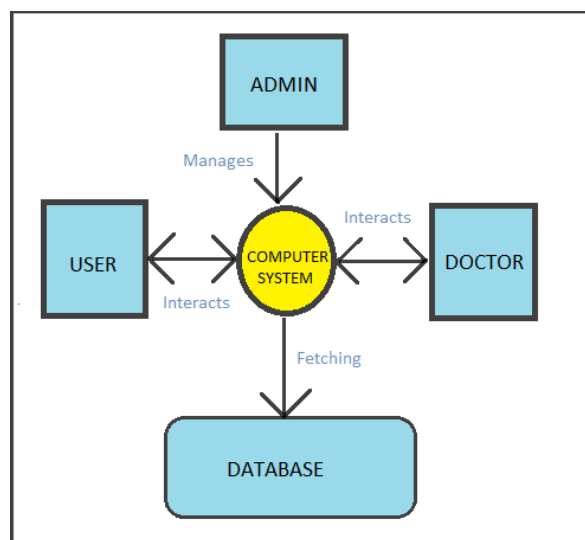


Fig.3 System Interaction overview diagram

## V. CONCLUSION

In this paper we have reviewed the applicability of Bayes algorithm for the prediction of disease from its respective symptoms. From data mining, classification is used and from training data set the classification rules are used to help the system in correct prediction. Users are able to get the easy medical diagnosis from installed application in their computers. It will also help the users to get the help or advice from specialist doctors through posting questions on their modules. While doctors can provide the trustworthy answers to patients, they can also discuss among themselves on various health issues. Users can use it irrespective of financial aspects. This system can be used both in urban and rural areas. It can provide the various healthy diet and routines to various patients and also to other normal persons. Our approach is to provide the user a reliable and secure application for the health prediction and medical diagnosis anytime and anywhere.

While some biometric techniques also can be used along with the system in future like heart beat measurement and blood pressure measurement. It can also be used along with other hardware modules to provide the whole home treatment to patients. In future the system can also provide the video calling among doctor and patients. This system will help the users in getting the normal health issues solved easily and will provide the users healthy life style in efficient manner.

## REFERENCES

- [1] "Online health research eclipsing patient-doctor conversations," Makovsky Health and Kelton, Survey, 2013
- [2] Disease Inference from Health-Related Questions via Sparse Deep Learning. Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Member, IEEE, Bo Zhang, Senior Member, IEEE, Tat-Seng Chua, Senior Member, IEEE. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING MAY 2014
- [3] Mining Health Examination Records — A Graph-based Approach. Ling Chen, Xue Li, Member, IEEE, Quan Z. Sheng, Member, IEEE, Wen-Chih Peng, Member, IEEE, John Bennett, Hsiao-Yun Hu, and Nicole Huang. IEEE TRANSACTIONS ON KNOWLEDGE DISCOVERY AND ENGINEERING
- [4] METEOR: An Enterprise Health Informatics Environment to Support Evidence-based Medicine. Mamta Puppala, Tiancheng He, Shenyi Chen, Richard Ogunti, Xiaohui Yu, Fuhai Li, Robert Jackson, and Stephen T. C. Wong. IEEE Transactions on Biomedical Engineering
- [5] G. J. Simon, P. J. Caraballo, T. M. Therneau, S. S. Cha, M. R. Castro, and P. W. Li, "Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus," IEEE Transactions Knowledge and Data Engineerin
- [6] <http://www.webmd.com/>
- [7] <https://www.healthtap.com>