

Health Prediction Using Data Mining

Richa Gupta, Poonam Bharti, Rashmi Sinha

Department of Information Technology, IIMT College of Engineering, Greater Noida, Uttar Pradesh, India

ABSTRACT

The paper presents an overview of the data mining techniques with its applications, medical and educational aspects of Clinical Predictions. In medical and health care areas, due to regulations and due to the availability of computers a large amount of data is becoming available. Such a large amount of data cannot be processed by humans in a short time to make diagnosis, and treatment schedules. A major objective is to evaluate data mining techniques in clinical and health care applications to develop accurate decisions. It also gives a detailed discussion of medical data techniques which can improve various aspects of clinical predictions. It is a new powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different databases to transform it into new research and results. It makes use of machine learning and database management to extract new patterns from large data sets and the knowledge associated with these patterns. The actual task is to extract data by automatic or semi-automatic means. The different parameters included in data mining include clustering, forecasting, path analysis and predictive analysis. The objective of this work entitled “Health prediction using data mining” is to provide a user friendly and easily understandable GUI to users to easily get instant guidance on their health issues through an intelligent health care system online. The main objective of the system is to predict disease according to symptoms and suggest list of doctors and medicines.

Keywords: Predict Disease, Health, Data mining, GUI

I. INTRODUCTION

It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason. The Health Prediction system is an end user support and online consultation project. Here we propose a system that allows users to get instant guidance on their health issues through an intelligent health care system online. The system is fed with various symptoms and the disease/illness associated with those systems. The system allows user to share their symptoms and issues.

It then processes user’s symptoms to check for various illnesses that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient’s symptoms.

“Health prediction system “ is the computerization of medical information to support and optimize

- (1) Administration of health services
- (2) Clinical care
- (3) Medical research
- (4) Training

It is the application of computing and communication technologies to optimize health information processing by collection, storage, effective retrieval. The proposed system is mainly used by all the people where confidentiality and integrity of the data has utmost importance.

In doctor module when doctor login to the system doctor can view his patient details and the report of that patient. Doctor can view details about the patient search what patient searched for according to their prediction. Doctor can view his personal details. Admin can add new disease details by specifying the type and symptoms of the disease into the database. Based on the name of the disease and symptom the data mining algorithm works. Admin can view various disease and symptoms stored in database. This system will provide proper guidance when the user specifies the symptoms of his illness.

This project is basically based on data analysis using data mining which is a trend in today's time. Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. The health prediction system is an end user support and online consultation project. Here we propose a system that allows users to get instant guidance on their health issues through an intelligent health care system.

II. BACKGROUND

The core idea behind this work is to propose a system that allows users to get instant guidance on their health issues. This system is fed with various symptoms and the disease/illness associated with those systems. This system allows user to share their symptoms and issues. It then processes user's symptoms to check for various illness that could be associated with it. If the system is not able to provide suitable results, it informs the user about the type of

disease or disorder it feels user's symptoms are associated with and suggest the doctor to whom he or she can contact.

To beat the downside of existing framework we have created health prediction system. We have built up a specialist framework called health prediction framework, which is utilized for improving initial level and proposes the possible diseases. It begins with getting some information about manifestations to the patient, in the event that the framework can distinguish the fitting sickness, at that point it proposes a specialist accessible to the patient in the closest conceivable territory.

On the off chance that the framework isn't sufficiently sure, it asks few questions to the patients, still on the off chance that the framework isn't sure; at that point it will show a few tests to the patient. In light of accessible total data, the framework will demonstrate the result. Here we utilize some intelligent methods to figure the most precise disorder that could be associated with patient's appearances and dependent on the database of a couple of patients restorative record, calculation (Naïve Bayes) is connected for mapping the side effects with conceivable diseases. This framework improves undertaking of the specialists as well as helps the patients by giving vital help at soonest organize conceivable.

III. SIGNIFICANCE OF THE STUDY

There is always a need of a system that will provide the disease information according to symptoms shared by user. This system will help the user to find good doctors and medicines. Data mining used in the field of medical application can exploit the hidden patterns in voluminous medical data which otherwise is left undiscovered.

The main goals of this work are:

- User friendly interface to provide ease.
- To get instant guidance on health issues.
- To get online medical facilities.

- Avoid problems while unavailability of doctors.
- User can search for doctor’s help at any point of time.
- User can talk about their illness and get instant diagnosis.
- Informs the user about the type of disease or disorder it feels.
- Doctors get more clients online.

The project work has a wide scope, as it is not intended to a particular organization. This project is going to develop generic software, which can be applied by any business’s organization. Moreover, it provides facility to its users. Also the software is going to provide a huge amount of summary data.

Hypothesis

Everyone is a patient at some time or another, and we all want good medical care. We assume that doctors are all medical experts and that there is good research behind all their decisions. However, that cannot always be the case. Nevertheless, they cannot possibly commit to memory all the knowledge they need for every situation, and they probably do not have it readily available. Disadvantage of an existing system would be that the patients have to visit the doctor in person and still does not get proper treatment, as the doctors are unable to predict exact disease.

- To overcome the drawback of existing system we have developed health prediction system. This system not only simplifies task of the doctors but also helps the patients by providing necessary help at an earliest stage possible.

IV. METHODS AND MATERIAL

- Existing system- Everyone is a patient at some time or another, and we all want good medical care. We assume that doctors are all medical experts and that there is good research behind all their decisions. However , that cannot always be the case. Nevertheless, they cannot possibly commit to memory all the knowledge

they need for every situation, and they probably do not have it readily available.

Even if they did have access to the massive amounts of data needed to compare treatment outcomes for all the disease they encounter, they would still need time and expertise to analyse that information and integrate it with the patient’s own medical profile. But this kind of in-depth research and statistical analysis is beyond the scope of a physician’s work. They want a doctor who will talk to them, listen to what they say and give them advice about how to get better and protect their health in the future. In many cases, the wish for a prescription is secondary to the wish of being cared for.

Disadvantage of an existing system would be that the patients have to visit the doctor in person and still does not get proper treatment, as the doctors are unable to predict the exact disease. Human error can be avoided with the help of computer assisted quality decision making. It is poor when there are huge amounts of data to be classified. In addition, efficiency and accuracy of decisions will decrease when humans are put into stress and immense work. Imagine a doctor who has to examine five patient records; he or she will go through them with ease. However if the number of records increases with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analysed.

V. RESEARCH METHODOLOGY

We have studied disease of many areas such as:

LUNGS:

DISEASES	SYMPTOMS
Idiopathic Pulmonary fibrosis (a disease in which tissues in your lungs	Shortness of breath, A dry, hacking cough,Rapid breathing,Gradual,

becomes thick and stiff or scarred overtime).	unintended weight loss, Tiredness, Clubbing, which is the widening and rounding of the tips of the fingers or toes.
Influenza (flu)	Fever, Dry persistent, Cough, Fatigue and weakness, Nasal congestion, Sore throat
Lung cancer	A cough that does not go away or gets worse, Fever. Chest pain, Hoarseness, Weight loss and loss of appetite, Coughing up blood or rust-colored spit
Pertussis (whooping cough)	Runny nose, Nasal congestion, Red, watery eyes, Fever Cough

LIVER:

DISEASE	SYMPTOMS
Acute liver failure	Yellowish of your skin and eye balls (jaundice), Pain in your upper right abdomen, Abdominal swelling, Nausea, Vomiting, A general sense of feeling unwell (malaise), Disorientation and confusion. Sleepiness.
Cirrhosis	Loss of appetite, Lack of energy which may be debilitating, Weight loss or sudden weight gain. Bruises, Yellowing of skin or the whites of eyes (jaundice), Itchy skin, Fluid retention (edema) and swelling in the ankles, legs and abdomen, A brownish and orange tint of the wine light colored stools, Confusion disorientation, personality changes. Blood in the stool, Fever.
Non-alcoholic fatty disease	Fatigue. Pain in upper right abdomen. Weight loss.
Alagille syndrome	Jaundice

VI. SAMPLING TECHNIQUE

CHEST:

DISEASES	SYMPTOMS
Heart attack	Chest pain, Sweating, Pressure Fullness or tightness in your chest, Crushing or searing pain radiating to your back, neck, jaws, shoulders and arms particularly left arms. Shortness of breath, Dizziness or weakness, Nausea or vomiting

1. Naïve Bayes algorithm

Naïve bayes algorithm is a classification algorithm based on Bayes' theorems use in predictive modelling and this algorithm uses Bayesian techniques. This algorithm is less computationally intense than other and therefore is useful for quickly generating mining models to discover relationships between input columns and predictable columns.

Data required for naïve bayes models

Requirements for a Naive Bayes model

1. A single key column – Each model must contain one numeric or text column that uniquely identifies each record. Compound keys are not allowed.
2. Input columns – In a Naive Bayes model, all columns must be either discrete or discretized columns it is also important to ensure that the input attributes are independent of each other.
3. At least one predictable column- The predictable attribute must contain discrete or discretized values. The values of the predictable column can be treated as inputs.
4. Viewing the model- To explore the model we can use the Microsoft Naïve Bayes Viewer. The viewer shows you how the input attributes related.
5. Making predictions- After the model has been trained, the results are stored as a set of patterns, which we use to make predictions.

We can create queries to return predictions about how new data relates to the predictable attribute.

VII.RESULT AND DISCUSSION

- Patient Registration – If Patient is a new user, he will enter his personal details, he will get a user Id, and password through which he can login to the system.
- Patient Login – Patient Login to the system using his ID and Password.
- Disease Prediction – Patient will specify the symptoms caused due to his illness. The system will ask certain symptoms regarding his illness and then predict the disease based on the symptoms specified by patient
- Search Doctor – Patient can search for doctor when the disease is predicted by the system which will provide him with the address of any nearby doctor.
- Get Appointment – Patient will choose an appointment date for the selected doctor.

VIII. FUTURE SCOPE

Concealed learning will be extracted from the verifiable information in the proposed framework, by getting ready datasets by applying apriori calculation. Anticipating savvy wellbeing should be possible just if framework reacts that way. These datasets will be contrasted and the approaching questions and the last report will be produced utilizing Association Rule Mining. Since this proposed system will chip away at genuine chronicled information, it will give exact and productive outcomes, which will enable patients, to get the conclusion in a split second. More work should be possible later on by utilizing more informational index identified with heart sickness and by utilizing diverse information decrease techniques to improve the characterization.

For better precision and expectation of heart sicknesses the datasets that will be used must be quality organized and free from special cases, inconsistencies, and missing characteristics. This web application can be additionally upgraded in an Android application. This will be accessible to clients on versatile premise and its utilization can be additionally expanded. Likewise, highlight like getting the specialist online on a visit with the goal that patients can straightforwardly converse with the concerned specialists. The modules doing malignant growth examination can be coordinated to discover how close the individual related with disease is. This will make this web application unsurprising in obvious sense.

This web application can be further enhanced in an android app. This will be available to users on mobile basis and its use can be further increased. Also, feature like getting the doctor online on chat so that patients can directly talk to the concerned doctors. The modules doing cancer analysis can be integrated to find how close the person associated with cancer is.

IX. CONCLUSIONS

- Data mining can be beneficial in the field of medical domain. However privacy, security and unable to log into account are the big problems if they are not addressed and resolved properly. It describes about the proposal of hybrid data mining model to extract classification knowledge for aid of various disease in clinical decision system and presents a framework of the tool various tools used for analysis.
- Sometimes the situation occurs when you need the doctor's help immediately, but they are not available due to some reason. In our project , we have designed a health prediction system, which is an online system and various patients from any location can view it. Our system comprises of main components such as patient login, enter symptoms in the system, and prescribe medicines, suggested nearby doctor. 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. Sometimes the situation occurs when you need the doctor's help immediately, but they are not available due to some reason. Thus , it allows the users to get analysis on the symptoms they give for predicting the disease they are suffering from.
- Data mining can be helpful in the field of restorative space. Anyway protection , security and unfit to sign into the record are the huge issues on the off chance that they are not tended to and settled appropriately. It portrays the proposition of a crossover information mining model to separate arrangements learning for the guide of different maladies in the clinical choice framework and presents a structure of the apparatus different devices

utilized for investigation. Now and again the circumstance happens when you need the specialist's assistance promptly, however they are not accessible because of some reason. In our venture, we have planned another wellbeing forecast framework, which is an online framework and different patients from any areas can see it.

- Our framework involves fundamental parts , for example, patient login, enter side effects in the system and recommend medications, proposes an adjacent specialist. The application takes the contribution of different manifestations from the patient, does the examination of the entered side effects and gives fitting sickness expectation. Our framework enables the clients to get an examination of the indications they give for anticipating the malady they are experiencing. Some of the time the circumstance happens when you need the specialist's assistance quickly, yet they are not accessible because of some reason. Along these lines, it enables the clients to get an examination of the side effects they give for anticipating the infection they are experiencing.

X. LIMITATIONS OF THE STUDY

- Supports the use of predictive model markup language(PMML) to create mining models.
- Supports drill through.
- Does not support the creation of data mining dimensions.
- Supports the use of OLAP mining models.

XI. REFERENCES

- [1]. For python installation, <https://www.python.org>
- [2]. For python basics, www.w3schools.com
- [3]. www.javatpoint.com
- [4]. <https://www.geeksforgeeks.org/python-diango/>

- [5]. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5228051&newsearch=true&queryText=health%20data%20mining>
- [6]. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7344518&newsearch=true&queryText=health%20data%20mining>