

# A Smart Health Prediction Using Data Mining

Prof. Krishna Kumar Tripathi<sup>1</sup>, Shubham Jawadwar<sup>2</sup>, Siddhesh Murudkar<sup>3</sup>, Prince Mishra<sup>4</sup>

<sup>1</sup>Professor, Dept. of Computer Engineering, Shivajirao S. Jondhale College Of Engineering, Maharashtra, India

<sup>2,3,4</sup> Student, COMP, Shivajirao S. Jondhale College Of Engineering, Maharashtra, India

\*\*\*

**Abstract** - This project can be used for the data mining techniques such as medical field, research field, and educational field and various aspects. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. As per the modern technology huge improvement has been made in computer field and therefore there is no need to deal with such a large amount of data at a same time. A major objective of this paper is to evaluate data mining techniques in clinical and health care applications to develop an accurate decisions. It is a famous and 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 researches and results. It makes use of Artificial Intelligence, machine learning and database management techniques to extract new patterns from large data sets and the knowledge associated with these patterns. By using this technique data can be extracted automatically or semi automatically. The different parameters included in data mining include clustering, forecasting, path analysis and predictive analysis.

**Key Words:** Health Prediction, Medicines, Doctors Appointment, Symptoms, Location.

## 1. INTRODUCTION

The health industry has been growing a lot from past few years. This technique has gained a lot of importance in medical areas. It has been calculated that a care hospital may generate five terabytes of data in the year. In our day to day life we have lot of other problems to deal with and we neglect our health problems. So in order to overcome such problem we have designed user friendly website which helps users to get diagnosed from their residence at any time. We also provide an option for booking an appointment with the doctor to discuss health related problems and get diagnosed properly. Data mining has variety of scopes in major fields some of which are listed below:

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

To analyze large amount of data, data mining technique is used. For each subfield of Clinical Predictions, and also presented how clinical data warehousing in combination with data mining can help administrative, clinical, research and educational aspects of Clinical Predictions. There are

various challenges of data mining which can be discussed in health prediction.

## 2. LITERATURE SURVEY

Divya Jain et.al presents a review of the implementation of Apriori Algorithm on datasets using machine learning tool Weka. Ruijuan Hu states the details of the idea on two-step frequent data items using Apriori algorithms and Association Rules. This mentions a new improvised method called Improved Apriori Algorithm to eliminate cons of Apriori algorithm. Gitanjali J, et.al proposed study of huge datasets from various angles and obtaining gist of useful information. These methods are useful in detecting diseases and providing proper remedy for the same. Krishnaiah et.al aims to calculate various methods of data mining in applications to develop decisions and also to provide a detailed discussion about medical. Data mining techniques can improve various angles of clinical predictions. Dan A. Simovici proposed that association rules represent knowledge in data sets as results and are directly related to calculation of frequent item sets. Mohammed Abdul Khaleel states data mining as a concept that studies large amount of data and extracts patterns that can be converted to useful knowledge.

In this paper, we set out to identify efficient algorithm for mining results. We can create versatile applications for medicine sector so as to fulfil by using all these predictive analytics and data mining techniques.

1. This tells how Naïve Bayes algorithm is used to find frequent data items and compares them with the existing algorithms.
2. Data mining techniques is used to apply on medical data which has abundant scope for improving health solutions.
3. Electronic health records and other historical medical data can prove miracles if used for a right purpose.
4. Huge amounts of complex data generated by health care sector includes details about diseases, patients, diagnosis methods, electronic patient's details hospitals resources.

## 3. PROPOSED SYSTEM

Sometimes the situation occurs when you need the doctor's help immediately, but they are not available due to some reason. This system allows the users to get analysis on the symptoms they give for predicting the disease they are

suffering from. User will be asked to enter the symptoms, then system will processes those symptoms for various illness or disease that user could be aliked with. In this system we use some techniques of data mining to guess the most accurate diseases or illness that could be related with patient’s symptoms.

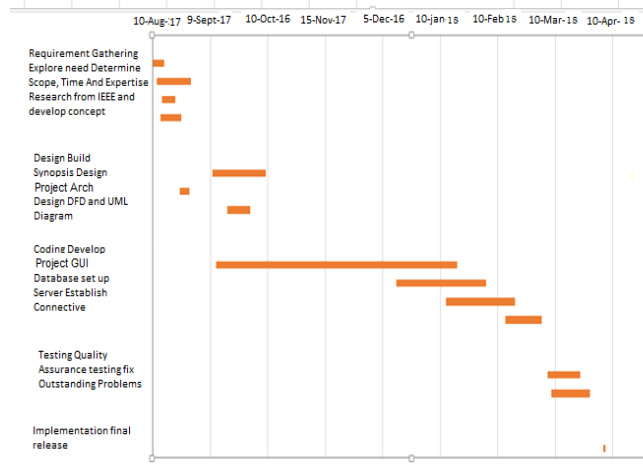


Chart -1: Gantt chart

If the system is unable to provide solutions, it informs the user about probable disease they have. If the user symptoms does not exactly match with any disease in our database, shows the diseases user could probably have judging the input symptoms. This system tends to replace the existing system for going to the doctor for getting diagnosis on illness you are suffering from to a smart solution where you get instant diagnosis on entering symptoms in the system.



Fig -2: Screenshot

The main features of this system will be giving instant diagnosis on the user entered symptoms and getting tips for remaining fit. In the proposed system, we will use the data mining method in which the symptoms entered by users are cross checked in the database and from that the frequent item sets are mined out of that the existing datasets. The proposed system is very efficient algorithm implemented.

#### 4. ALGORITHM USED

##### 4.1 Naïve Byes

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 naïve byes classifiers. For example we can say that the fruit is an apple if it is a red, round and about the 10cm in diameter. The naïve byes algorithm also considers the each features to identify the fruit is apple.

There are many probability models, out of which for some types the naïve byes algorithm is efficiently trained in supervised learning training.

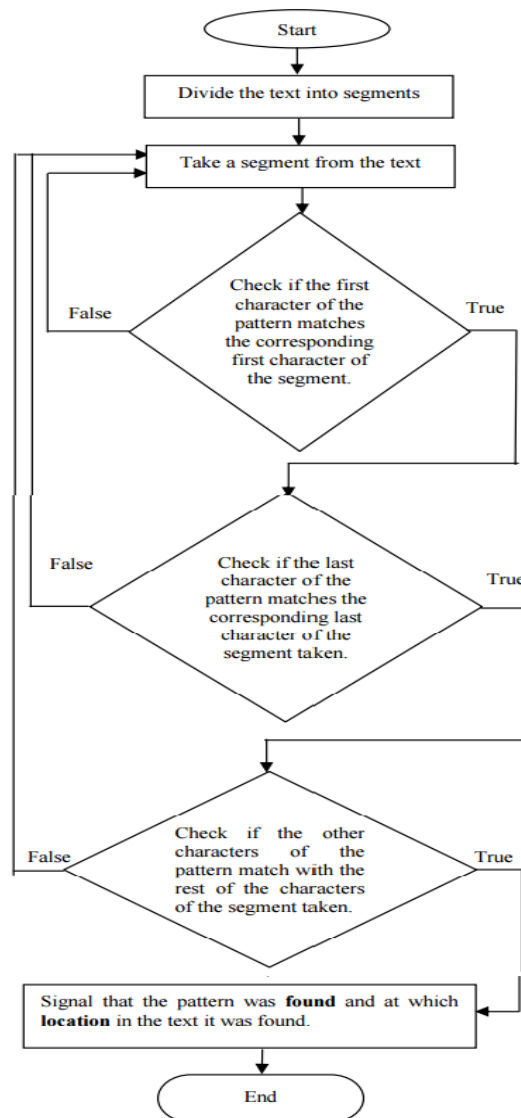


Fig -3: Naive Byes Flowchart

## 4.2 Naive Bays Algorithm Steps

### Step 1:

First we will divide the text into the segments in that the first segment starts from the element which is presented at index 0, the second segment start from the element which is presented at index 1 and so on. The segments which has taken has shifted one character from previous one.

### Step 2:

After that we will compare the first character of that pattern with the first character of corresponding segment taken, if there is a match occurs, then we will go to the next step. And if the mismatch occurs, then we will take the next segment and we will repeat step 2.

### Step 3:

After that compare the last character of that pattern with the last character of corresponding segment taken, if there is a match occurs, then we will go to the next step, and if a mismatch occurs, then we will take the next segment and go to step 2.

### Step 4:

Perform character by character comparison for the rest of the characters of the pattern with the rest of the characters of the segment taken (that is without considering the first and last characters in the comparison). If a mismatch is encountered while matching in any step of the comparison, then we stop comparing and proceed with the next segment for comparison and go to step 2, else continue comparing. If all the characters of the pattern match with the characters of the segment then signal that the pattern was found and at which location in the text it was found. Then we will proceed with the next segment and repeat step 2, for searching other occurrences of the patterns.

## 5. RESOURCE USED

### 5.1 Hardware:

1. Processor : Minimum 2.0GHz requires.
2. Ram : 2 GB.
3. Hard Disk : 100 GB.
4. Input device : Standard Keyboard and Mouse.
5. Output device: VGA and High Resolution Monitor.

### 5.2 Software:

1. Operating System : Windows 7.
2. Language : Java
3. Database : MySQL 5.0 & Above.
4. Tool : JDK 1.5 & Above, Eclipse IDE.
5. Server : Tomcat

## 6. APPLICATIONS

There are various application of data mining such as telecommunication industry, commercial industry, biological data analysis and many more.

With the advance changes happening in the technology, especially in the field of health organization a lot of data is produced day by day. Since there is need of analysis of data and the amount of data analyzed is in large amount, so there is need of excessive knowledge regarding the technology of data mining.

For health diagnosis E-healthcare applies data mining and telecommunication techniques. E-health was primarily used for patient data analysis and disease diagnosis at various levels. There are some patients who require continuous checkup and might need doctor help immediately.

## 7. CONCLUSION AND FUTURE WORK

In the proposed system, hidden knowledge will be extracted from the historical data by preparing datasets by applying Naïve Byes algorithm. Predicting smart health can be done only is system responds that way. These datasets will be compared with the incoming queries and the final report will be generated using Association Rule Mining. Since this proposed methodology will work on real historical data, it will provide accurate and efficient results, which will help patients get diagnosis instantly. This system will also guide the users of how to remain healthy and fit using tips provided here. The further enhancements that can be done would be integrating this web application 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 persson associated with cancer is. This will make this web application predictable in true sense.

## ACKNOWLEDGEMENT

We express our deep gratitude to our project guide (Prof. Krishna Kumar Tripathi) for providing timely assistant to our query and guidance that he gave owing to his experience in this field for past many year. He had indeed been a lighthouse for us in this journey. We extend our gratitude to our project coordinator Prof. Uttara Gogate, our Head of Department Prof. P. R. Rodge and our Principal Dr. J. W. Bakal.

We extend our sincerity appreciation to all our Professors from SHIVAJIRAO S. JONDHALE COLLEGE OF ENGINEERING for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual.

## REFERENCES

- [1] Mohammed, Zaki. J, Meira, Wagner, "Data Mining and Analysis: Fundamental Concepts and Algorithm", 2014.
- [2] Holzinger, Andreas, Rocker, Carsten, Ziefle, Martina, "Smart Health", 2015
- [3] Longhi, Sauro, Freddi, Allesandro, "Human Monitoring, Smart Health And Assisted Living", 2017.
- [4] Koehler, "Smart Health: What Today's Doctors Aren't Telling You", 2016
- [5] Thareja, Reema, "Data Warehousing", 2009.
- [6] Mannila, Heikki, Hand, David. J, Smyth, Padhraic, "Principles of Data Mining", 2001.
- [7] Dean, Jared, "Big Data, Data Mining and Machine Learning", 2014.