INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)

Available online at www.ijariit.com

Health Expert System – Prediction of Disease using Data Mining

Vinayak Sharma
vinayak.sharma@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Gayatri Mulchandani
gayatri.mulchandani@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Shivani Subnani
shivani.subnani@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Gaurav Gianani
gaurav.gianani@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Anjali Yeole
anjali.yeole@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

ABSTRACT

The main purpose of data mining application in health care system is to develop automated tool for identifying and disseminating relevant healthcare information. The objective of our system to provide prediction of disease depending on symptoms so as to take proactive treatment against disease. The system would reduce the human effort, reduce cost and time constraint in terms of human resources and expertise, and increase the diagnostic accuracy. In most developing countries, insufficiency of medical specialist has increased the mortality of patients suffering from various diseases. Insufficiency of medical specialists will never be overcome in short period of time. The main idea of project is to assist doctors, who fail to detect fatal diseases. The intelligent doctor will accept symptoms of patient. The symptoms, and the databases are matched to produce list of diseases and sufferings with their probabilities. We have used Apriori and Frequent Pattern Growth algorithm for predicting the disease for a given set symptoms. The whole process can be termed as KDD. [1]

Keywords: *Disease Prediction, KDD (Knowledge Discovery in Databases and Data Mining), Apriori Algorithm, Frequent Pattern Growth Algorithm.*

1. INTRODUCTION

Every human being faces health related issues each & every day where one cannot come to know what kind of disease/health problems are been faced. To identify what is the cause depending on the symptoms you can come to know what kind of disease you have & what treatment is required. Certain issues related to resources, lack of knowledge in medical science such kind of process may help doctors as well as patient to cure their disease by prediction analysis. Detecting diseases at early stages can enable to overcome dangers. However, waiting for students to become doctors and doctors to become specialists, many patients may already die. Current practice for medical treatment required patients to consult specialist for further diagnosis and treatment. Other medical practitioner may not have enough expertise or experience to deal with high risk diseases. As most of the high risk diseases could only be cured at the earlier stage, the patients may have to suffer for the rest of life. The AI is a study to evaluate human intelligence into computer technology. It starts with asking about symptoms to the patient, performs computations and provides the possible diseases. The system which is fully trained with knowledge & can display disease depending of what kind of symptom an individual possess. Also with such system one can keep track of their health.

2. KNOWLEDGE DISCOVERY IN DATABASES AND DATA MINING

KDD is the process of changing the low- level data into high-level knowledge. KDD includes statistics, database systems, computer programming, machine learning, and artificial intelligence. The Knowledge Discovery in Databases process comprise of a few steps

leading from raw data collections to some form of new information. The iterative process consists of the following steps: [1]

- *Data cleaning:* Also known as data cleansing it is a phase in which noise data and unrelated data are removed from the collection.
- *Data integration:* At this stage, several data sources, often heterogeneous, may be shared in a common source.
- *Data selection:* At this step, the data related to the analysis is decided on and retrieve from the data collection.
- *Data transformation:* Also known as data consolidation, it is a phase in which the chosen data is transformed into forms appropriate for the mining procedure.
- *Data mining:* It is the essential step in which clever techniques are applied to extract patterns potentially useful.
- *Pattern evaluation:* This step, firmly interesting patterns representing knowledge are known based on given measures.
- *Knowledge representation:* It is the last phase in which the discovered knowledge is visually represented to the user.

3. HEALTHCARE DATA MINING

Like analytics and business intelligence, the term data mining can mean different things to different people. The most basic definition of data mining is the analysis of large data sets to discover patterns and use those patterns to forecast or predict the likelihood of future events. We generally categorize analytics as follows:

- *Descriptive analytics*— Describing what has happened
- *Predictive analytics*— Predicting what will happen
- Prescriptive analytics— Determining what to do about it

It is to the middle category—predictive analytics— that data mining applies. Data mining involves uncovering patterns from vast data stores and using that information to build predictive models.

4. METHDOLOGY

The core objective of project is to develop a android application using data mining concept accompanied by Python 3.0 and MySQL. The whole process can be termed as "knowledge discovery process, (KDD)". This is because here we need to predict the disease for user input symptoms where the predicted disease is in the form of information or knowledge. Following Figure shows the steps carried out to predict the probable disease for inputted patient symptoms:

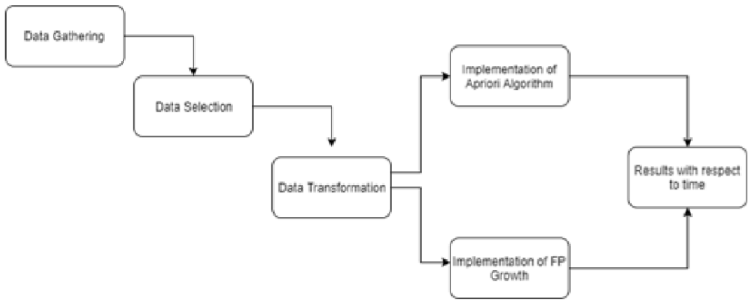

Fig 1: Steps Used Carried Out for Prediction

Preprocessing

Data gathering and selection: The raw data has been collected from the World Wide Web and data relevant to our purpose has been selected for further processing.

Data Transformation

After the first step the data is transformed into xls file so as to form a standard database. This xls file is extracted or read using file handling concept and stored in a database using MySQL. From the xls file only those information are read which are associated with the basic objective of our intended application. For example various stops words, verbs and adjective irrelevant to the application are kept behind and only the key meanings are read from the xls file so that it becomes easier for the application to implement the algorithms i.e Apriori and FP Growth over the disease-symptom database. We have used Apriori and FP Growth algorithm for predicting the disease for a given set symptoms. These symptoms are provides by an user as inputs. On accepting these inputs the application executes these algorithms over them by accessing the database created using MySQL in step 2 during preprocessing stage.

A. APRIORI ALGORITHM

The Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Apriori is a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and group of candidates are tested against the data). Apriori is designed to operate on database containing transactions, (for example: collection of items bought by customers).Key Concepts are:

- Frequent Itemset Search: Obtain item occurrence:
  Items that occur more than one times in the entire dataset.
  Get frequent item sets.

Generate items that occur frequently.

- Apriori property: Any subset of frequent item set must be frequent.
- Obtain rules that have greater confidence: Rules which satisfy minimum confidence are listed.

Algorithm Apriori
// Mn: Item set of size N
// Fn : frequent item set of size N
Step 1: F1 = {frequent items};
Step 2: for (N = 1; N<=Fn; N++) {
Step 3: Mn+1 = Medical symptoms derived from
Fn after entering first symptom
Step 4: each t transaction in the database do {
Step 5: Increment count of all the symptoms in Mn+1
Step 6: Fn+1= min_support medical data in Mn+1
}
Step 7: end}
Step 8: return union NFn;

Here we have implemented the Apriori algorithm by generating only one candidate set. This is because here our motive is to predict only one disease for a set of inputted symptoms.

B. FP GROWTH ALGORITHM

FP Growth stands for frequent pattern growth. It is a scalable technique for mining frequent patterns in a database. FP Growth is a two-step procedure. *Step1:* Build a compact data structure called the FP-Tree. (Build using two passes over the data set). *Step 2:* Extracts frequent item sets directly from the FP-Tree.

FP tree Generation: Item sets are considered in order of their descending value of support count. To facilitate tree traversal, an item header table is built so that each item points to its occurrences in the tree via a chain of node links. Frequent item generation: Frequent items are directly extracted from the FP tree on the basis of maximum frequency.

5. FLOW CHART

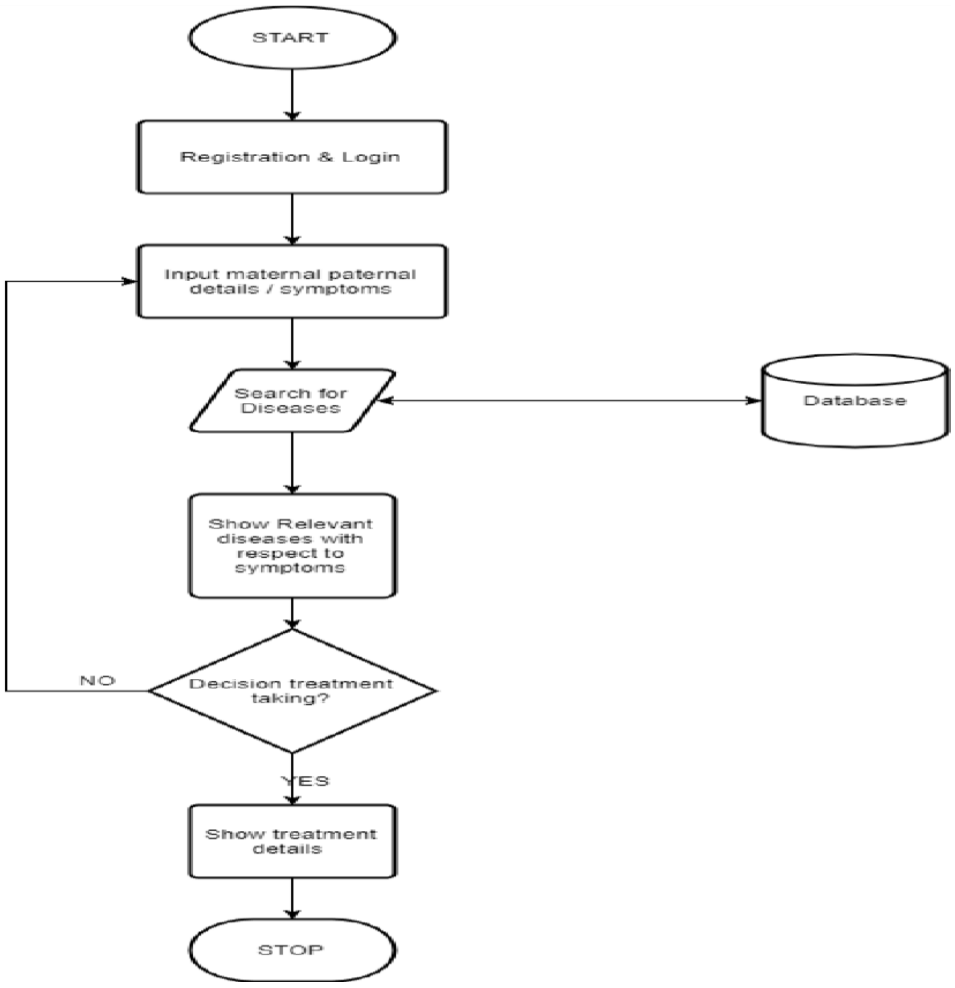

Fig 2: Flow Chart

6. RESULTS AND DISCUSSIONS

The result of our system will consist of the diseases and its respective accuracy level the patient is currently suffering from. Accuracy level of a specific disease will be based on the patient's age, gender, his own medical history and his maternal and paternal history. The result will assist the clinical doctors, and the doctors will finally be able to cure the patient depending on the diseases that have higher accuracy.

7. CONCLUSION

In this paper, we've proposed a system that aims at reducing mortality of patients suffering from various diseases. This system will be of benefit to the doctors by making the diagnosis more reasonable. Here, we have presented a model with an efficient approach considering multiple parameters. Data sets are provided so that system can learn and map the symptoms.

8. REFERENCES

[1] Dr.B.Srinivasan, K.Pavya (2016), "A Study on Data Mining Prediction Techniques in HealthCare Sector", International Research Journal of Engineering and Technology (IRJET), Vol. 3(3)

[2] S.Vijayarani, M. Divya, "An Efficient Algorithm for Generating Classification Rules",
IJCST, vol. 2, Issue 4, 2011

[3] Muhamad Hariz Muhamad Adnan,Wahidah
Husain, Nur'Aini Abdul Rashid (2012), "Data
Mining for Medical Systems: A Review", International conferences on advances in computer and information technology.

[4] K. Rajalakshmi & Dr. S. S. Dhenakaran(2015)," Analysis of Datamining Prediction Techniques in Healthcare Management System", International Journal of Advanced Research in Computer Science and Software Engineering,Vol.5,Issue4.

[5] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[6] RubanD.Canlas Jr., MSIT. MBA, — Data mining in Healthcare: Current applications and issues.

[7] HianChyeKoh and Gerald Tan, —Data Mining Applications in Healthcare, journal of Healthcare Information Management – Vol 19, No 2.

