International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

Disease Prediction and Doctor Recommendation System

Dhanashri Gujar[1], Rashmi Biyani[2], Tejaswini Bramhane[3], Snehal Bhosale[4], Tejaswita P. Vaidya[5]

[1,2,3,4] B.E. (Computer Engineering), Sinhgad College of Engineering, Pune, Maharashtra, India
[5]Assistant Professor, Dept. of Computer Engineering, Sinhgad College of Engineering, Pune, Maharashtra, India

---

Abstract - *Properly analyzing clinical documents about patients' health anticipate the possibility of occurrence of various diseases. In addition, acquiring information regarding specialists of that particular disease as per the requirement facilitates proper and efficient diagnosis. This paper provides a novel method that uses data mining technique, namely, Naïve Bayes classification algorithm for prediction of disease followed by recommendation of specialists of the predicted disease. Using medical profiles such as heart rate, blood pressure through sensors and other externally observable symptoms such as fever, cold, headache etc. that patient has, prediction of likelihood of a disease is done. Naïve Bayes algorithm takes these symptoms and predicts disease. Furthermore, all the needful and adequate information regarding the predicted disease as well as the recommended doctors is provided. Recommendation suggests the location, contact and other necessary details of the disease specialists based on the filters chosen by the user out of less fees, more experience, nearest location and feedback reviews of the doctors. Reviews are compared using Stanford's CoreNLP algorithm. Thus user can get appropriate treatment and necessary medical advice as fast as possible. Additionally, users provide their feedback for the recommended doctors which are then added for analysis in order to make further recommendations based on reviews.*

***Key Words*: CoreNLP, Naïve Bayes, Prediction, Recommendation, Weka.**

1. Introduction

Healthcare industry generates terabytes of data every year. The medical documents maintained are a pool of information regarding patients. The task of extracting useful information or quality healthcare is tricky and important [6]. By analyzing these voluminous data we can predict the occurrence of the disease and safe guard people. Thus, an intelligent system for disease prediction plays a major role in controlling the disease and maintaining the good health status for people by providing accurate and trustworthy disease risk prediction.

1.1 Prediction System

In this paper, the focus is on data mining techniques to extract hidden rules and relationships between symptoms and diseases. Disease prediction is done by medical profiles

such as blood sugar, blood pressure, blood oxygen, headache and other symptoms. Based on this, the most probable disease is predicted by Naive Bayes classifier.

1.2 Recommendation System

In addition to this, the specialists for the predicted disease are recommended based on filters chosen by user. Review based recommendation is done by fetching the reviews of various doctors provided by the previous users. This is followed by feeding these to CoreNLP for processing. Location based recommendation gives the nearest specialist relative to user's current location. Fees based and experience based recommendation consider user's preferences among various ranges.

2. Motivation

At present in order to remain healthy, regular body diagnosis is necessary. Today, there are multiple sources available as individual prediction or recommendation system but the need of the hour is to have an integrated model comprising both. Also, it would be more appropriate and convenient if people could get basic diagnosis online 24x7 rather than visiting hospitals & clinics frequently. Thus, reducing cost and saving time. If certain anomalies found in the diagnosis then recommendation of nearby specialist and hospitals according to user's preference would facilitate in quick and appropriate treatment. Healthcare being a domain evolving continuously and generating a huge amount of data develops a need to use the data for useful knowledge which attracts large organizations to invest heavily in this field.

3. Related Work

Binal T. et al [1], Healthcare decision support system for swine flu prediction using naïve bayes classifier, focuses on the aspect of medical diagnosis by learning patterns through the collected data for swine flu using naïve bayes classifier for classifying the patients of swine flu into three categories (least possible, probable or most probable), resulting into an accuracy of nearly 63.33%. Datasets used for this classification were limited in number.

Shengyong W. et al [2], Predicting Disease by using data mining based on healthcare information system, describes the experiments of applying data mining to disease prediction from a large number of real world medical records of hypertension. This paper compares three

algorithms– naïve bayes, J-48 and ensemble of five J-48 classifiers. Here, naïve bayes & J-48 showed nearly same accuracy of 83%.

Marcelo M. et al [3], A collaborative filtering approach based on user's reviews, proposes a collaborative filtering

approach that uses users' reviews to produce items description. The reviews are processed using CoreNLP tool and then the algorithm creates a representation which is used to compute similarity of items which is used in collaborative filtering approach based on k-nearest neighbors.

Manjusha K. et al [5], published Prediction of different dermatological conditions using Naive Bayesian classification, gives the possibilities of eight diseases using patient's attributes. The system extracts hidden knowledge from the database. System can also predict diseases other than dermatological diseases.

4. Proposed Work

An intelligent system for accurate disease prediction and medical facilities recommendation plays a major role in effective treatment. This system takes the symptoms from users and predicts the most accurate disease accordingly. Additionally, sensor module helps in continuous evaluation of vitals like heart rate, blood pressure and sugar level for patient which is fed in the system at runtime for analysis along with other external symptoms. Based on the prediction, system recommends the hospitals/clinics according to user's preference out of nearest location, less fess, more experience and better reviews with doctors having expertise for that particular disease to avail the required medications. Also, the users can provide their feedback for the recommended doctors.

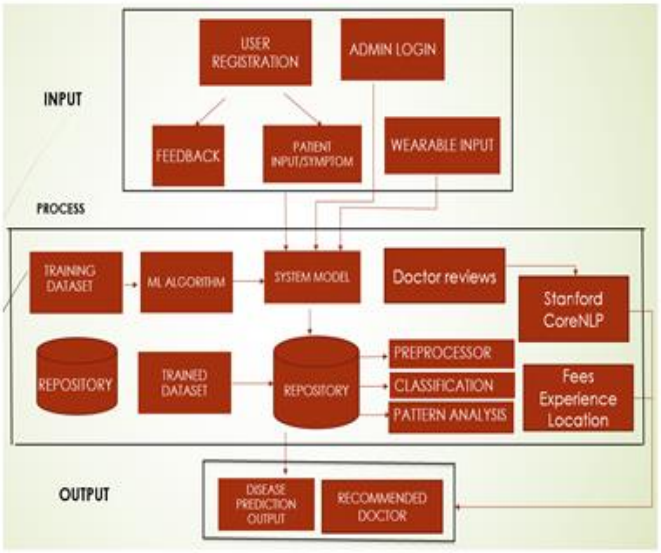

Figure 1: System Overview Diagram

5. Implementation Techniques

5.1 Naïve Bayes

A Naïve Bayesian classifier, is a model joint probability distribution over a set of stochastic variables. Instances of the classification problem under study are presented to the classifier as a combination of values for the feature variables; the classifier then returns a posterior probability distribution over the class variable. Learning such a classifier amounts to establishing the prior probabilities of the different classes and estimating the conditional probabilities of the various features given each of the classes [1].

According to Bayes theorem of probability theory:-

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad (1)$$

It is assumed that attributes E_1 to E_m are class conditionally independent, which means it is often assumed that

$$P(E|H) = \sum_{i=1}^{N} P(Ei = vi|H) \quad (2)$$

After making the above assumption, the classifier is called Naïve Bayes classifier.

Table 1: Input Attributes

Gender: F/M
Blood Sugar: 72-162 mg/dl
Blood Pressure: 120/80 mmHg
Pulse Rate: 60-100 bpm (normal)
Fever: Yes/No
Level of Fever: low(1), moderate(2), high(3)
Cough, Cold: <7 days =1, >7 days =2
Pain: chest, muscles, body, abdominal, no pain
Breathlessness: Yes/No
Headache: normal=1, severe=2
Vomiting, Weakness, Chills, Constipation, Dizziness, Loss of Appetite: Yes/No

5.2 Weka

Naive Bayes has been implemented using weka libraries. Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling. It supports several data mining tasks such as data preprocessing, classification, clustering etc.

The accuracy, performance measures and confusion matrix for Naive Bayes when applied on training dataset on Weka platform are shown below:

```
=== Summary ===

Correctly Classified Instances         275               92.2819 %
Incorrectly Classified Instances        23                7.7181 %
Kappa statistic                          0.8969
Mean absolute error                      0.0413
Root mean squared error                  0.1734
Relative absolute error                 11.0427 %
Root relative squared error             40.0856 %
Total Number of Instances              298

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.986    0.009    0.972      0.986   0.979      0.972    0.998     0.994     Typhoid
               0.886    0.044    0.861      0.886   0.873      0.834    0.983     0.961     pneumonia
               0.877    0.044    0.865      0.877   0.871      0.828    0.982     0.948     influenza
               0.941    0.005    0.988      0.941   0.964      0.950    0.998     0.995     hypertension
Weighted Avg.  0.923    0.025    0.924      0.923   0.923      0.898    0.991     0.975

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 69  0  1  0 |  a = Typhoid
  0 62  7  1 |  b = pneumonia
  2  7 64  0 |  c = influenza
  0  3  2 80 |  d = hypertension
```

Figure 2: Statistics for Naïve Bayes on Weka

5.3 CoreNLP

CoreNLP provides a set of natural language analysis tools written in Java for text processing. It takes raw English text input and generates a complete structured analysis of the most common NLP routines. CoreNLP is an integrated framework for several language analysis tools, called annotators [3]. Some of the relevant annotators comprised in this tool are: tokenizer, part-of-speech (POS) tagger, sentence splitter and the sentiment analysis tool.

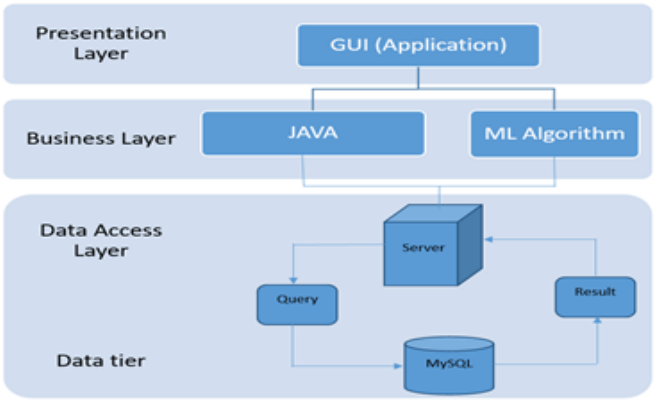

Figure 3: Architecture Diagram

6. Results

The most probable disease predicted from the probabilities calculated by the Naïve Bayes classifier with an expected accuracy of above 80 percent. Recommendation of top five specialists corresponding to the predicted disease and filters chosen by user.

7. Conclusion and Future Work

Application of data mining techniques for disease prediction from a large number of real world medical records and CoreNLP techniques for doctor recommendation from the reviews of previous users has been studied and implemented.

The system can further be improved by incorporating various other symptoms and increasing the number of cases for training and testing. Additionally, considering the calorie count, step count, sleep quality and other medical profile through wearable, required nutritional diet plan can be suggested. Also, alerts and notifications can be sent timely to the user as well as his/her guardian in case of any risks.

References

[1] Binal A. Thakkar, Mosin I. Hasan, Mansi A. Desai, "Healthcare decision support system for swine flu prediction using naïve bayes classifier", "IEEE", 101-105, 2010.

[2] Feixiang Huang, Shengyong Wang, Chien Chung Chan, "Predicting disease by using data mining based on healthcare information system", "IEEE", 2012.

[3] Rafael M. D'addio, Marcelo G. Manzato, "A collaborative filtering approach based on users' reviews","IEEE", 204-209, 2014.

[4] F. O. Isinkaye, Y.O. Fola Jimi, B.A. Ojokoh," Recmmendation systems : Principles, Methods and Evaluation", "Elsevier", 261-273,2015.

[5] Manjusha K.K.,K. Sankaranarayanan, Seena P,"Prediction of different dermatological conditions using naïve Bayesian classification","IJARCSSE",864-868,2014.

[6] Subhash C. Pandey, "Data Mining techniques for medical data: A Review", "IEEE", 2016.

[7] Lee, Yunkyoung, "Recommendation system using collaborative filtering", 2015.

[8] Geoff A., Elvis C., Mickey E., Sameul S., "Smart wearable body sensors for patients self assistant and monitoring", BioMedCentral, 2014.

[9] G Gannu, Y Kakodkar, A Marian, "Improving the quality of predictions using textual information in online users reviews", Information Systems, 2013.

[10] Sellappan P., Rafia A., "Intelligent heart disease prediction system using data mining techniques", IEEE/ACS, 2008.