Risk Factor Of Cancer Diabetic And Heart Disease On Naive Based Algorithm

Dr. T. Arumuga Maria Devi,B.E.,M.Tech.,Ph.D.Miss. S. Sowmiya, M.Sc, PG Scholer,

Asst Prof,*CITE, MS University,Tirunelveli, India* *CITE, MS University,Tirunelveli, India*

***Abstract*— This project, makes a survey about different classification techniques used for predicting the those 3 disease of each person based on age, gender, Blood pressure, cholesterol, pulse rate. It is implemented using naïve based algorithm. It retrieves hidden data from stored database and compares the user values with trained data set. The application answer, complex queries for diagnosing the disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot and finally gives accurate result and find out the disease problem for the particular persons. To enhance visualization and ease of interpretation, it displays the results in tabular and PDF forms.**

INTRODUCTION

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease." Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help overcome this situation.

I. LITERATURE REVIEW

Analysis of Medical Data using Data Mining and Formal Concept Analysis,

Anamika Gupta, Naveen Kumar, and Vasudha Bhatnagar IEEE2005

This paper applies the techniques of classification in data mining and context reduction in Formal concept Analysis on medical data and finds out the redundancies among the medical examination tests prescribed for diagnosis of a disease. Currently the technique works on binary data, which means that a test can have either positive result or negative result. In future we will experiment on quantitative values of the test, which means working on multi-valued context.

Disadvantage: uses complex medical dataset as input

"Multi-class classification algorithm for optical diagnosis of cancer",

A Guptha and S Guptha, December 2005

This projectreport development of a direct multi-class spectroscopic diagnostic algorithm for discrimination of high-grade cancerous tissue sites from low-grade as well as precancerous and normal squamous tissue sites of human oral cavity. A multi-class diagnostic algorithm based on TPCR has been developed and used for classification of auto fluorescence spectral data acquired from the oral cavity of patients screened for neoplasm of oral cavity into four classes – normal, leukoplakik, low-grade SCC and high-grade SCC.

Disadvantage: doesn't involve prediction, only one disease was analysed, high cost due to equipment cost

II. IMPLEMENTATION

I will try to analyze the risk factoe of diseasefor a person I will develop the code in java step by step and see the practical implementation of risk analysis.

The code is divided into following parts:

.

1. Activity diagram
2. Sequence diagram
3. Collaboration diagram
4. Component diagram and Deployment diagram
5. Use case diagram

A. Activity diagram

It shows the activities that are carried out in the proposed system. The generated questionnaire according to the choice of the disease is sent to the client system. The input got from the client and the datasets will be used by the system which uses naïve bayes classifier to generate the chance of occurrence of the disease i.e., if the disease will occur or not. The generated result will be sent to the client system. If necessary it will be sent to the doctor's mail address for further verification.

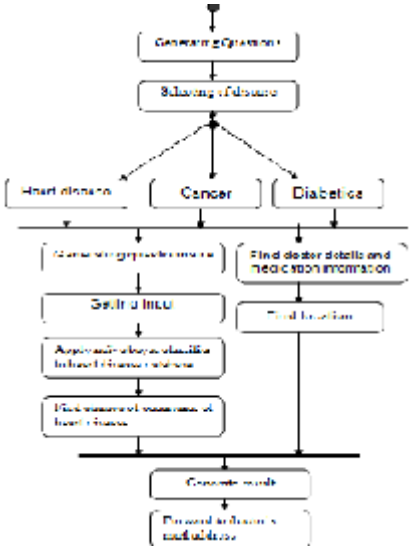

Figure 2.1– Activity diagram

B. Sequence diagram

The generated questionnaire according to the choice of the disease is sent to the client system. The input got from the client and the datasets will be used by the system which uses naïve bayes classifier to generate the chance of occurrence of the disease. The generated result will be sent to the client system. If necessary it will be sent to the doctor's mail address. This data can also be used to store in the database for future reference by the doctor.

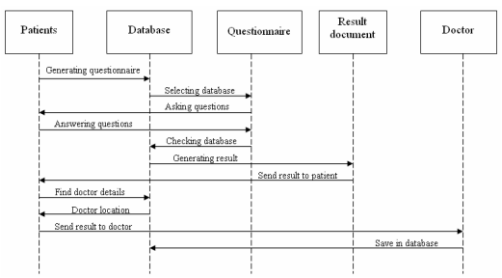

Figure 2.2 – Sequence diagram

C. Component and Deployment diagram

The component diagram shows how the physical components or packages interact with one another. The above diagram shows that the client and the server are related to one another, using the input given by the user and the output generated. Deployment diagram shows the configuration of runtime processing elements and software components, processes and objects that live in them.

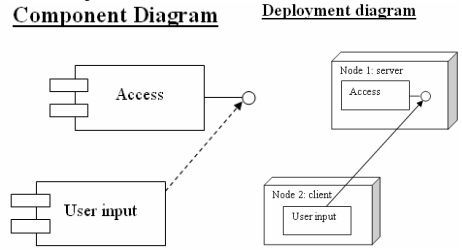

Figure 2.3 Component and Deployment diagram

E. Collaboration diagram

The type of disease is first chosen by the client. The corresponding questionnaire is sent to the client system. The input got from the client and the datasets will be used by the system which uses naïve bayes classifier to generate the chance of occurrence of the disease. The generated result will be sent to the client system. If necessary it will be sent to the doctor's mail address. This data can also be used to store in the database for future reference by the doctor.

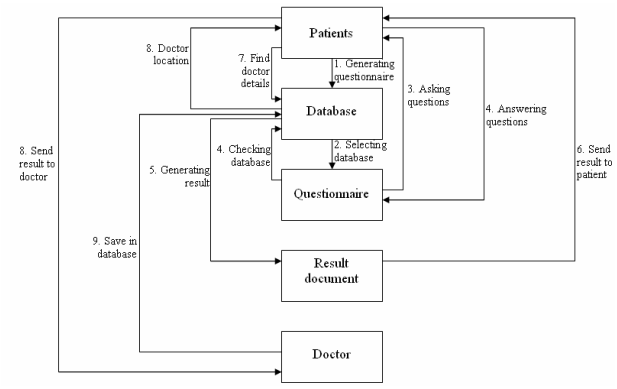

Figure 3.4 Collaboration diagram

F. Use case diagram

The database is available on the server side. The type of disease is chosen by the client. The corresponding questionnaire is sent to the client system. The input got from the client and the datasets will be used by the system which uses naïve bayes classifier to generate the chance of occurrence of the disease. The generated result will be sent to the client system. If necessary it will be sent to the doctor's mail address. This data can also be used to store in the database for future reference by the doctor.

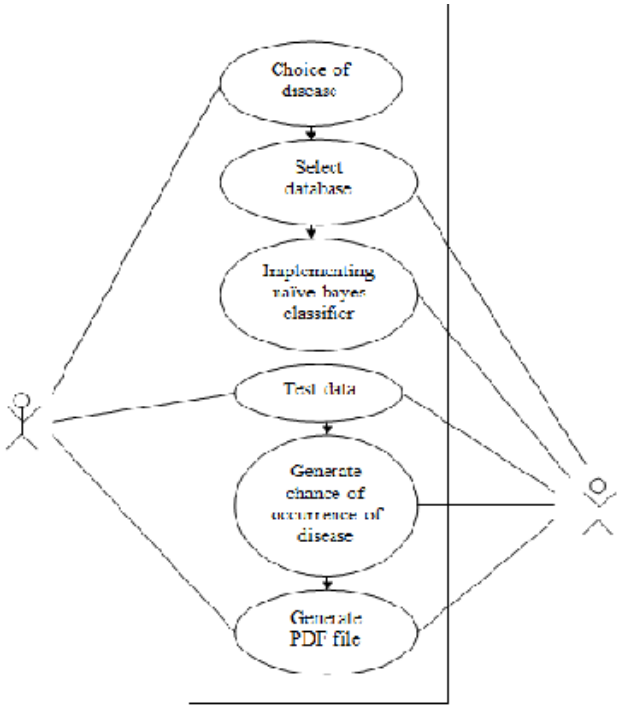

Figure 2.5 Use case diagram

III. EXISTING PROBLEM

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease." Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. IHDPS can be further enhanced and expanded. For example, it can incorporate other medical attributes besides the 15 listed and suggestions such as doctor details and medications to patients could also be provided.

PROPOSED SYSTEM

The proposed system uses data mining technique "Naïve Bayes classifier" for the construction of the prediction system. It can predict three diseases which are diabetes, cancer andheart attack. This system involves higher number of data sets and attributes which are directly collected from doctor's information for accurate prediction of the disease. Nearly 300 records for each disease will be collected and stored in the database. 16 medical attributes for heart disease, 9 medical attributes for diabetics and 10 medical attributes for cancer are taken. These datasets will be used for prediction of the heart disease using data mining technique. Since more attributes and more data sets are available, the datasets can be used to predict more accurate occurrence of the disease. This causes the disease to be predicted more effectively.

Moreover this proposed system also consists of various suggestions such as doctor details and prescriptions. Each disease will have different specialists for analyzing the disease. The details of each doctor along with their location for each disease will be given. Cost of visiting the doctor in the initial stage could be avoided since the medications will be prescribed.

E. Methods of Naïve Bayes Algorithm

1) Analyzing the Data set

A **data set** (or **dataset**) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the data set in question. It lists values for each of the variables, such as height and weight of an object or values of random numbers. Each value is known as a datum.The attribute "Diagnosis" was identified as the predictable attribute with value "1" for patients with heart disease and value "0" for patients with no heart disease. The attribute "PatientID" was used as the key; the rest are input attributes. It is assumed that problems such as missing data, inconsistent data, and duplicate data have all been resolved

2) NaivesBaye's Implementation in Mining

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and A represents the prior event. Bayes' Theorem

Prob (B given A) = Prob (A and B)/Prob (A)

3) Designing the Questionnaire for Heart, Cancer and Diabetes

Questionnaires have advantages over some other types of medical symptoms that they are cheap, do not require as much effort from the questioner as verbal or telephone surveys, and often have standardized answers that make it simple to compile data. However, such standardized answers may frustrate users. Questionnaires are also sharply limited by the fact that respondents must be able to read the questions and respond to them.

4) Diagnosis In WEB system

In our Heart disease development the modeling and the standardized notations allow to express complex ideas in a precise way, facilitating the communication among the project participants that generally have different technical and cultural knowledge.MVC architecture has had wide acceptance for corporation software development. It plans to divide the system in three different layers that are in charge of interface control logic and data access, this facilitates the maintenance and evolution of systems according to the independence of the present classes in each layer. With the purpose of illustrating a Successful application built under MVC, in this work we introduce different phases of analysis, design and implementation of a database and web application..

F. BENEFITS

It can answer complex queries for diagnosing disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot .By providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation

IV. EXPERIMENTAL RESULTS

A. Start Servlet

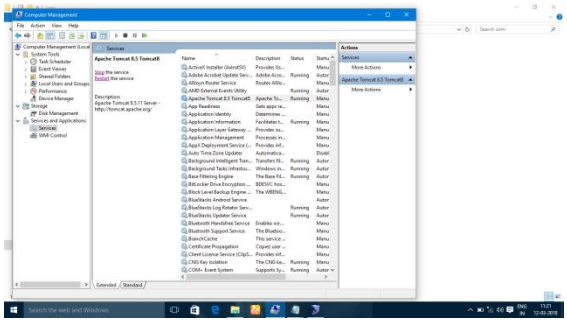

Figure 4.1 Start Servlet

B. Login page

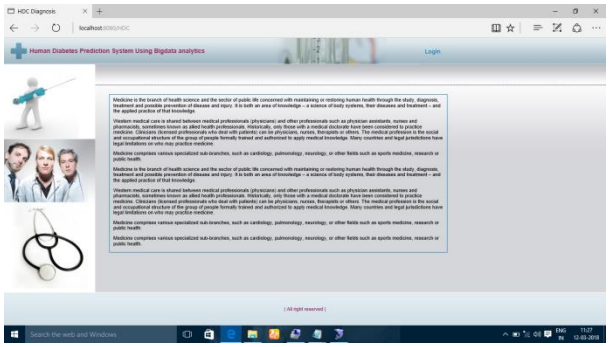

Figure 4.2 – Login page

C. Register form

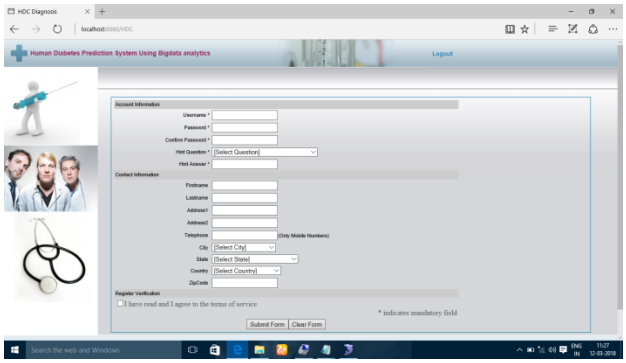

Figure 4.3 – Register page

D. Fact sheet

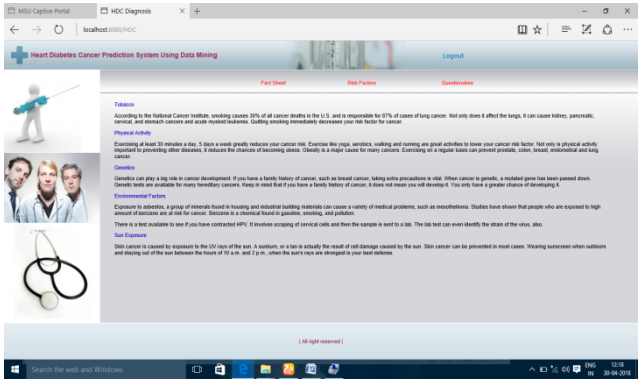

Figure 4.4 – Fact Sheet

E. Risk factor

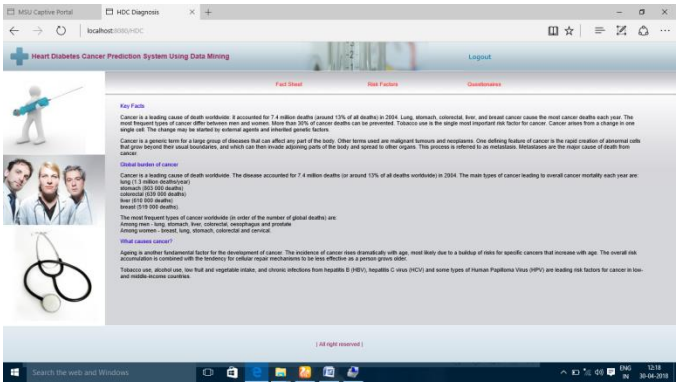

Figure 4.5 Risk Factor

F. Questioners

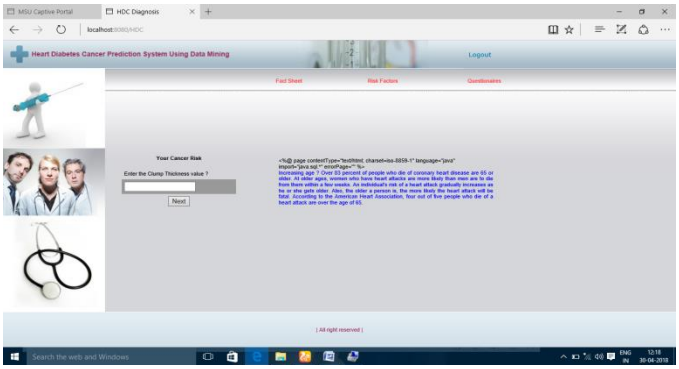

Figure 4.6 Questionari

G. *Saving details*

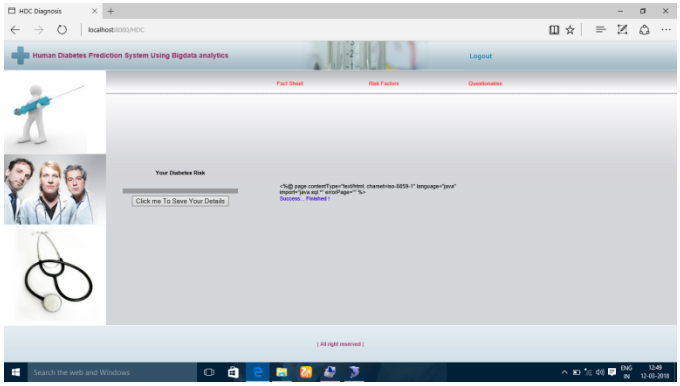

Figure 4.7 Saving detail

H. *View report*

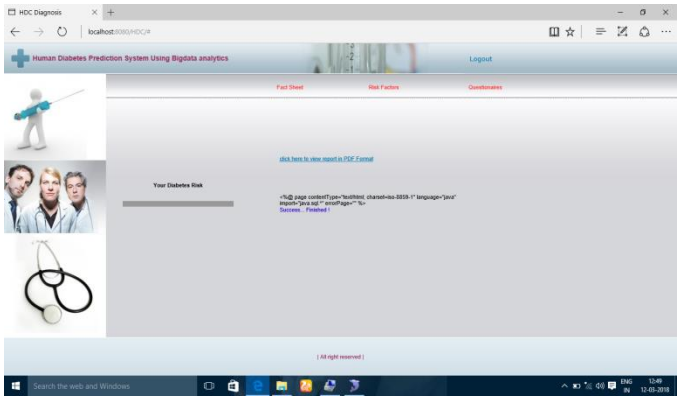

Figure 4.8 View report

V. CONCLUSION

A. *CONCLUSION*

In this paper Naïve based techniques is used to build a disease risk prediction system is proposed. Heart disease, diabetic, Cancer has become the leading cause of death world-wide. The most effective way to reduce deaths is to detect it earlier. Many people avoid disease screening due to the cost involved in taking several tests for diagnosis. This prediction system may provide easy and a cost effective way for screening disease and may play a pivotal role in earlier diagnosis process for different types of disease and provide effective preventive strategy. This system can also be used as a source of record with detailed patient history in hospitals as well as help doctors to concentrate on particular therapy for any patient.

.

B. *FUTURE ENHANCEMENT*

1) *Multi-Class Classification*

Till now, I have only dealt with binary classification of tweets, either as positive or negative sentiment. There are many tweets, for instance, those with URL's which do not have any sentiment, or, are neutral. These tweets are mainly for sharing some useful information with people, and not necessarily for raising an opinion.

2) *More Numeric Features*

The numeric features that were used in this experiment include number of negative and positive words, emoticons, length of tweets and number of special characters such as exclamations, hash tags and so on. The numeric features did not yield good accuracy and gave around 63 percent accuracy. Hence, as a part of my future work on this, I would like to generate more as well as smarter numeric features

.

3) *Use More Classifiers*

In this project Natural Language Processing, Stop Words were used extensively. I would also like to explore other machine learning algorithms like Artificial Neural networks. Also generation of more numeric features will allow me to use more binary classifiers such as logistic regression and so on.

REFERENCES

1. Ali Bharami, “Object oriented Systems development”, Tata McGraw-Hill Edition, New Delhi, pp.103-114
2. Florin Gorunescu, "Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection," International Journal of Bioengineering, Biotechnology and Nanotechnology, Vol.1, No.2, pp.105 - 108, 2008.
3. Roger S. Pressman, “Software engineering”, sixth edition, pp.226-230, McGraw-Hill international edition, 2006.
4. Mehmed, K.: “Data mining: Concepts, Models, Methods and Algorithms”, New Jersey: John Wiley, 2003.
5. Tang, Z. H., MacLennan, and J.: “Data Mining with SQL Server 2005”, Indianapolis: Wiley, 2005.
6. https://www.researchgate.net/publication/312188365_Heart_Attack_prediction_using_Data_mining_technique.
7. https://www.researchgate.net/publication/303524487
8. https://www.futuremedicalsights.com/reports/big-data-mining-in-disease-overview-and-analysis
9. https://www.researchgate.net/.../236966138_DATA_MINING_FOR_CANCER
10. https://www.researchgate.net/journals.sagepub.com/doi/abs/10.1177/0047287517696960
11. https://www.tandfonline.com/doi/full/10.1080/13683500.2012.71832
12. https://skemman.is/bitstream/1946/23184/1/FELMAN_lilja_rognvaldsdottir.pd

BIOGRAPHY

Dr. T. Arumuga Maria Devi Received B.E. Degree in Electronics & Communication Engineering from Manonmaniam Sundaranar University, Tirunelveli India in 2003, M.Tech degree in Computer & Information Technology from Manonmaniam Sundaranar University, Tirunelveli, India in 2005 and also received Ph.D Degree in Information Technology – Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli , India in 2012 and also the Assistant Professor of Centre for Information Technology and Engineering of Manonmaniam Sundaranar University since November 2005 onwards.

Ms. S. Sowmiya Received B.Sc. Degree in Information Technology from Manonmaniam Sundaranar University, Tirunelveli India in 2016, and Currently, doing M.Sc Information Technology in Centre for Information Technology and Engineering of Manonmaniam Sundaranar University, Tirunelveli, India.