IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

A Predictive Approach for Blood Distribution through Data Mining Techniques

Ms. Priyanka Sheshrao Nerkar[1], Prof. Leena K. Gautam[2]

M.E (Computer Engineering), Sant Gadge Baba University, Amravati, Maharashtra[1]

Computer Science & Technology, Sant Gadge Baba University, Amravati, Maharashtra[2]

Abstract: Blood is extremely important thing for any hospital. But getting it on a right time is one of the difficult task for any hospital. Many times lack of blood results into delay in emergency operation. This problem can be solved by predicting the need of blood for any hospital in advance. This can be achieved by using data mining technique. Data mining will help to predict need of blood in any hospital for any duration of time. Predicted information will help to distribute blood which is mainly concentrated at few places in hospital across a country. This entire infrastructure will help us to keep right amount of blood at right place. Traditionally, blood record is usually managed by blood banks and hospitals in the form of hand written documents manually. These blood banks are located across different areas/cities in a country. Telephone is the only way of communication for these blood banks. So, in the case of any urgent requirement, blood has to carry from one blood bank to another and this process takes hours or days. Ultimately, this results into delay in operations or emergency can be produced. Solution to above problem is a proposing framework in which record is collected from sources and stored in a system then by prediction system will determine which hospital or blood banks need how much amount of blood. So, this system will definitely remove the problems come in traditional way, By using this system need of blood will definitely can be predicted in advance and it will help in emergencies.

Keywords: Data mining, Prediction analysis, K Means algorithm, Bulk Data Entry, Single Data Entry, REST web services.

1. INTRODUCTION

The main agenda behind this research is improving blood distribution across large geometric area and reducing the emergency because of insufficient blood. These all things can be achieved by doing predictive analysis of blood requirement and data mining will help to do this analysis with great accuracy. This system can easily collect records from various sources. For collecting these record we have proposed two types of systems [1].

First one will be the Bulk Data processing system and another one will be Single data entry system. The main purpose behind this system is that accepting the records from various sources and send it into the system.

First type of design is the Bulk Data Processing [1]. Hospitals which maintains generally have their own system to maintain such records. Such hospitals will upload their date in a bulk into this system. Each entry get separated with comma and each entry gets processed[3].

Now, the second type of design will be using is the real time data entry and processing [4]. In this system targets the hospitals those don't maintain records in computerized format, instead they maintain records in handwritten documents like manual jounals. To convert that data into computerized format, to create one web base application. This application will be easy to use and ask basic details for each data entry [5]. Hospitals can add there records to this web application at any time, it may be daily or weekly or once in the month. just need to collect this data and upload this data to our system. In this way we can convert non-computerize data into computerize data format [6].

Below are the main topics which are going to cover in research:

1. Collecting blood records from various sources [7].
2. Storing data into database.
3. Pre-processing those records and make it available for in ready to use state [8](finding consumption on weekly basis).
4. Processing that data and finding the solution which is mainly, where to distribute the blood and what is the required quantity.
5. Re-updating the data after entire distribution.

This architecture is mainly divided into three phases.

1. Collecting the data from various data source and storing that data.
2. Processing the data and convert it into ready to use state and taking input the blood data from blood bank and calculating the need of blood for hospitals, i.e,final distribution result.
3. Once the result is ready then distribute the blood on the basis of that result.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

2. LITERATURE REVIEW

The paper published in Hawaii International Conference On System science by authors MadhavErraguntla, Peter Tomasulo, Kevin Land, HanyKamel, they have used JAVA SERVER PAGES(JSP) technology for Browser (client) and Server interaction.

2.1 JAVA SERVER PAGES (JSP):

JavaServer Pages (JSP) is a technology that helps software developers create dynamically generated web pages based on HTML, XML, or other document types. Released in 1999 by Sun Microsystems, JSP is similar to PHP and ASP, but it uses the Java programming language. Java Server Pages (JSP) is a technology that helps software developers create dynamically generated web pages based on HTML, XML, or other document types.
Released in1999 by Sun Microsystems, JSP is similar to PHP and ASP, but it uses the Java programming language.
JSP allows Java code and certain pre-defined actions to be interleaved with static web markup content, such as HTML, with the resulting page being compiled and executed on the server to deliver a document. The compiled pages, as well as any dependent Java libraries, contain Java bytecode rather than machine code. Like any other Java program, they must be executed within a Java virtual machine (JVM) that interacts with the server's host operating system to provide an abstract, platform-neutral environment. JSPs are usually used to deliver HTML and XML documents.
JSP has its own set of disadvantage. As java code is combine with html code. No separate business logic and presentation layer, Pages became massive. Java code scattered throughout making understanding what the page is doing a difficult effort due to the mixed models of Java and HTML. Also it has limited abilities to usefully re-use code across pages. As java code is mix with html code, there error handling is very difficult. All to gather maintaining and updating JSP is very difficult and complex job. In today's technology world REST architecture is more upgraded technology for client-serve interaction, which is used in this framework. REST architecture has more advantages over JSP.Details of REST services has been given in the paper.

3. PROPOSED WORK

This system is mainly divided into three phases. First is data collection phase and storage, second is data processing phase and generating result and third is final result distribution. Below is the detail description of each phase.

3.1 Data Collection and Storage:

3.1.1Bulk Data Entry and Single Data Entry:
Clients(Hospitals/Blood Banks) will be given a FORM through which they can upload their data to the system daily or weekly manner. This front end part which is visible to the clients can be created through scripting or programming languages like HTML/PHP/.NET etc. The data collected from clients need to store in database[6]. Before storing data in database in the form of table, some processing is required on raw information which is the after called as DATA. This processing is done through following technologies [4]:

Bulk Data Processing: Hospitals which maintains generally have their own system to maintain such records. Such hospitals will upload their date in a bulk into this system. Each entry get separated with comma and each entry gets processed[3].

Single Data Procesing: This technique used for the hospital which don't have data stored in computerized format. These hospitals maintain data manually [4]. Now, to convert that data to computerize format, one web base application is devlopes. This application will be easy to use and ask basic details for each data entry. Hospital can add there data to this web application at any time, it may be daily or weekly or once in the month. Now data is in required format so no need to process this data again and again. just need to collect this data and upload this data to our system[5].

3.1.2 Data Storage:
Data will store in the form of relational database management system (RDBMS).
As, the data is store in database, in order to transmit it to user interface it needs to convert it in appropriate format. The interaction between web base user interface and database is done with the help of middleware. This middleware use REST services to communicate with user interface. REST based communication will give us grate level of flexibility and modularity.This model organizes data into one or more tables (or "relations") of columns and rows, with a unique key identifying each row. Rows are also called records or tuples[7].

3.1.3 REST web services:
Representational state transfer (REST) or RESTful web services are one way of providing interoperability between computer systems on the Internet. REST-compliant web services allow requesting systems to access and manipulate textual representations of web resources using a uniform and predefined set of stateless operations. It is used for client server interaction between server and web browser (clients). It processes the data on a server with the help of Java Code ,sends output in appropriate format in XML etc.It does this because web browser understands only HTML statement and business logic is written in java code. So, to enable communication between these two this framework uses REST services.

REST services can be called via URL using java script. A web service is a collection of open protocols and standards used for exchanging data between applications or systems.Software applications written in various programming languages and running on various platforms

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

can use web services to exchange data over computer networks like the Internet in a manner similar to inter-process communication on a single computer. Web services based on REST Architecture are known as RESTful web services. These web services use HTTP methods to implement the concept of REST architecture. A RESTful web service usually defines a URI, Uniform Resource Identifier a service, provides resource representation such as JSON and set of HTTP Methods.

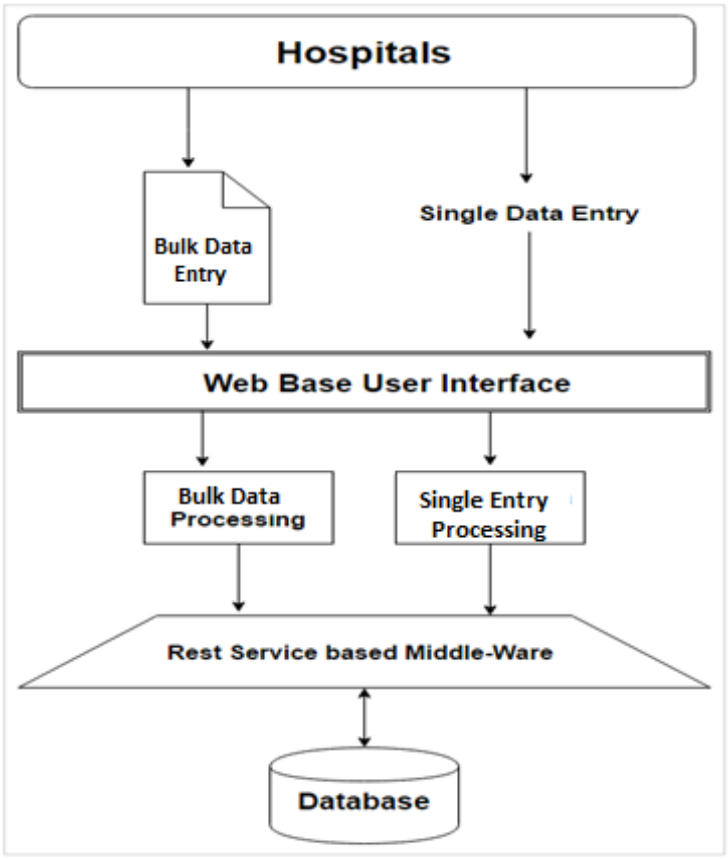

Fig.1: Components of System Architecture

3.1.4 System Architecture:

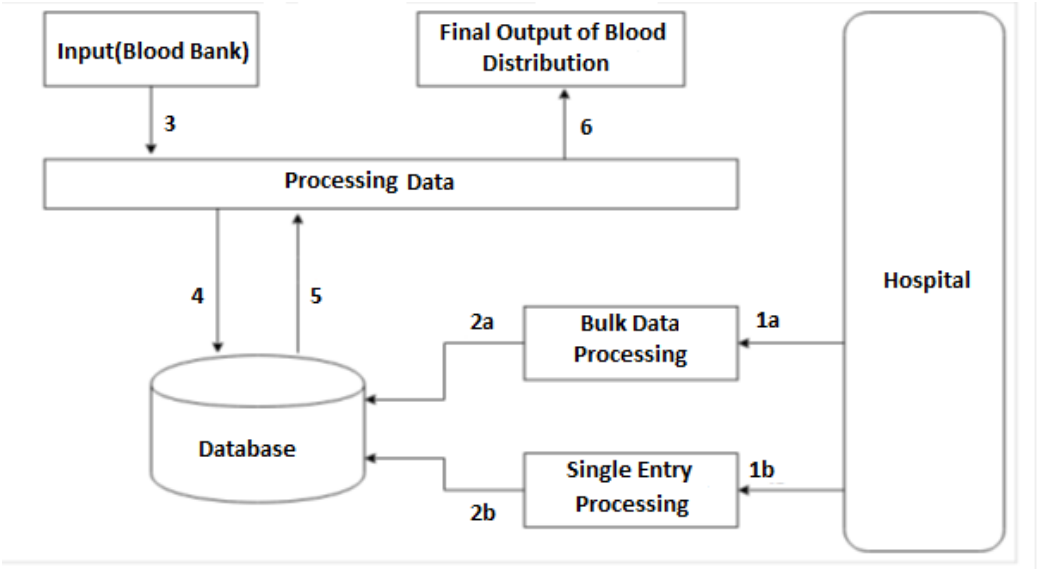

Fig. 2: Architecture of a System with Database

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

3.2 Processing the Data:

3.2.1 K-Mean algorithm: k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means because of the k in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by k-means to classify new data into the existing clusters.

In this framework ,K-Means algorithm is applied on consumption value of every week which is stored in 2D array. Algorithm takes any random numbers which acts as mean. In this system two random numbers are taken, this is because, if no. of clusters increases then inputs for further prediction will be less, inputs for calculating average will be less. The 2 clusters which are formed with the given input set, one will be of smaller consumption values and other will be of higher consumption value[15].

The algorithm steps for k Means are as below:

1. Array of blood consumption on weekly basis is created, and this acts as a input to form clusters.
2. First two values are considered as means M1 and M 2.
3. Each input value are compared with the two means,those values which are closer to mean 1 or mean 2 are arrange in clusters C1 and C2 respectively.
4. Recalculate the means. Based on new means,rearrange the input values for M1 and M2 in clusters C1 and C2.
5. Now, it compares newly formed clusters in step 5 and old clusters in step 3. And if they are not identical then it will repeat step 3 and step 4.unless,old and new clusters are identical.
6. Last two set of clusters are final output of k-mean.

3.2.2 Predictive Analysis:

All the record is stored in a database with the respective date. Queries are fired to extract the record with respective data and then do the analysis on those records[1]. Steps are as follows:

1. Extracts coming Sunday from first entry date.
2. Extracts end date from DB.
3. Calculates no. of days in between start date (first Sunday) and end date (last Sunday).
4. Calculates,no. of weeks = (No. of days)/7
5. Creates input array of size of no.of weeks.
6. Finds coming Saturday from starting date (sunday).
7. Finds the amount of blood consumption in between starting and end date.
8. The values from step 7, are assigned into the array in step 5. Each index of array acts as one week.
9. Calculates next day to temp end date(saturday).
10. Creates object to call method from K-mean class[15].

3.3 Final Distribution:

Once all the processing is done and predicted value is calculated, this result is given to blood banks. Here, the question is which blood bank will give amount of blood to which hospital?

Now, here the answer is blood banks are mapped to specific hospitals in advance in a database.

Hence, likewise blood is distributed to respective hospitals and re-updating the data after entire distribution.

4. APPLICATIONS

This is reusable architecture. Same architecture can be used to predict the amount of different kinds of resources required and to distribute those resources across large geometrical area to the various consumer. For example, sugarcane can be distributed very efficiently in various sugar factories across India by predicting which sugar factory requires how much amount of sugarcane i.e., considering the demand of sugar in those places. Similarly, vegetables or other natural resources can be distributed.

5. CONCLUSION

This entire logic is designed for efficient blood distribution across all over the country and avoid any delay in operation or any emergency case due to insufficient blood. More and more efficient data mining algorithm will help to predict the requirement of blood with great efficiency and great accuracy. To enhance this concept, system can be designed by considering natural disasters or man- made attacks by which system can be failed. This can be added a new system which considers all these factors. In order to create good system with great accuracy and efficiency, all these scenarios need to be considered and system should know how to react and handle in these cases. In this way the problems arises in traditional way can be solved by this framework. This is my proposal to solve this problem and research is still going to generate more accurate and efficient system.

REFERENCES

[1] Vanderbilt University Online document.—The most accurate way to calculate total blood volume, http://www.mc.vanderbilt.edu/documents/vmcpathology/files/TBV%20calucation.pdf.
[2] Y. Lin. Support vector machines and the bayes rule inclassification. Data Mining and Knowledge Discovery

[3] F. Provost and T. Fawcett. Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining.
[4] Eder, Anne F.; Notari IV, Edward P.; and Dodd, Roger Y. ―Do reactions after whole blood donation predict syncope on return donation?‖ TRANSFUSION Volume 52, December 2012.
[5] Erraguntla, M., Gopal, B., Ramachandran, S., Mayer, R. J. 2012, ―Inference of Missing ICD9 Codes Using Text Mining and Nearest Neighbor Techniques,‖ Hawaii International Conference on System Sciences, 2012.
[6] Erraguntla, M., Ramachandran, S., Wu, C., and Mayer, R. J. 2010, ―Avian Influenza Datamining Using Environment, Epidemiology, and Etiology Surveillance and Analysis Toolkit (E3SAT),‖ Hawaii International Conference on System Sciences, 2010.
[7] https://link.springer.com/article/10.1023%2FA%3A1009769707641?LI=true
[8] Wieling, Wouter; France, Christopher R.; Dijk, Nynke van; Kamel, Hany; Thijs, Roland D and Tomasulo, Peter ―Physiologic strategies to prevent fainting responses during or after whole blood donation,‖ TRANSFUSON, April 2011.
[9] Snehal A. Mulay, P. R. Devale, G.V. Garje.―Decision Tree based Support Vector Machine for Intrusion Detection,‖ 2010 International Conference on Networking and Information Technology.
[10] http://rest.elkstein.org/2008/02/what-is-rest.html.
[11] F. Provost and T. Fawcett. Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pages 43–48. AAAI Press, 1997.
[12] Kamel, Hany; Townsend, Mary; Schroeder, Kadi;Bravo, Marjorie; Erraguntla, Madhav; Whitaker, Barbee; Land, Kevin; Tomasulo, Peter. ―UNITSTATESDONORHEMOVIGILANCE SYSTEM: Experience of 3participating blood centers,‖ AABB Annual Meeting & CTTXPO 2011 , San Diego, CA.
[13] KBSI, Environment, Epidemiology, and Etiology Surveillance and Analysis Toolkit Phase II, 2009,W81XWH-08-C-0756.
[14] http://www.servicearchitecture.com/articles/webservices/representationalstatetransferres.thtml.
[15] https://en.wikipedia.org/wiki/K-means_clustering