To improve Blood Donation Process using Data Mining Techniques

NetrikaW. Dhoke, Prof. Swati S. Deshmukh

Student of M.E, Dept of CSE, Raisoni College of Engineering and Management, Amravati, India.

Assistant Professor, Dept of IT, Raisoni College of Engineering and Management, Amravati, India.

ABSTRACT: Blood is the vital thing for human being becausethere is no alternative for replacing it. This makes it necessary for the presence of a system for predicting continuous behavior of healthy blood donors in blood transfusion organization. The aim of predicting behaviorof donor's is to determine and collect proper information about the blood donor, his health status and blood groups for providing future continuity to blood bank. As the Urgent need of blood is rapidly growing it is desire need to find the blood donor information properly and efficiently. It is becoming more and more difficult to extract the information using the conventional database techniques. So, in this paper we proposed the solution for properly mining the proper and required donor information from large amount of blood donor's database. It is necessarytoanalyzethe blood donor database,along with his information from blood banks as it is useful when urgent need occurs. For retrieving the proper information from large database related to the blood donor's data mining is used to analyze their availability, number of donors and all related information. There are various different techniquesand algorithmspresent in data mining like classification, clustering, association, etc. which suits better for desired task.

KEYWORDS: Blood Donation, Classification, Centroids, Data Mining, K-means clustering, Knowledge Discovery from Databases (KDD).

I. INTRODUCTION

Blood is always in great demand from the past, but nowadays as the population grows the ratio of road accidents, disease and medical surgeries are also growingin same amount. The blood and there donors are very much important as it cannot be manufactured and only come from generous donors. This gives us the importance of looking for blood and providing conditions for donating blood [1]. Blood donation and its transfusion service is an indispensable part of medicine and healthcare sectors [2]. Blood donation is a series of interdependent operations in blood transfusion organization that consists of donor registration, donors' health evaluation, donors availability information, blood collection, blood screening, supply when needed, inventory management, and proper dissemination of blood[3].

Nowadays information and computer technology has gaining more importance in medicine and healthcare sectors, as it is needed to make it efficient by using computer Technology (CT) to database systems. As the medical information and Healthcare sectors repositories data is complex in the computer technology [3] it explores the use of data mining field. For gaining useful and important information of large health related data from large amount of blood banks to give proper and timely delivery of blood data mining is very much important field. The volume of data in the blood donation is much large, raw dataset are not applicable for mining it. It is necessary to analyze them and convert it into useful information or knowledge. Therefore, we use data mining tools and techniques for finding

useful information [5] and it provides automatic pattern recognition. The data mining techniques for uncovering patterns in data from large datasets which is difficult to detect with traditional statistical methods [4].

There are lots of measures of the blood donors that are important while taking his blood and in future when there is need of the same [2]. Despite these measures, various adverse events and reactions can and do occur during and after blood donation, which cannot be experience by donors. Blood transfusion organization have to perform the task to take care of donors and minimize any chances of reaction that patient suffer during blood donation process.Blooddonor's places responsibility on healthcare givers and the users of blood donations, to avoid wastage and unnecessary use of blood transfusions from any donor. As there are lots of factors to be considered while taking the blood from any donor [24].As its registration with all his personal information, last donation date and any reaction occurred, etc. These all stored information is possible to access with the different techniques of data mining like classification, clustering, prediction, association, etc. These techniques gives the segments of potential blood donors in terms of identifiable characteristics, behavior patterns, and it alsosuggest properly at the time of blood requirement in urgent cases.

As the number of accidents, medical surgeries and health diseases are growing on increasing, therefore urgent need for blood is also growing towards the alarming rate.It is becoming more and more difficult to extract the information using the conventional database techniques. So, in this paper we proposed the solution for properly mining the proper and required donor information from large amount of blood donor's database. Firstly by using the classification technique and any of its suitable algorithm like Naivey Bayes, J48 Decision tree, Random tree, etc. we can make the different classes of donors according to some criteria like age group, Location, Physical fitness, Gender, etc. After that Clustering analysis methods that helpfor identifying natural grouping of data objects from the large datasets so that it can be mine easily. Clustering method of data analysis can be called as unsupervised classification which partition a set of data objects into meaning full sub classes. We apply this clustering analysis [6] technique on that single class to get more specific information regarding that specific group of donor that can be called at any time directly based on their particular attributes that are suitable at any time and in case of emergencies also and can be done using various algorithm [6] here we are using the k-means clustering algorithm which gives the perfect distance measure. The K-means clustering gives the proper measure that is a need to come up with a solution for analysis and obtain the proper information of the blood donors [1].This paper focus on analyzing the efficiency and need of the blood donor's data and this can be properly get by classification and Clustering algorithm in data mining that we are explained further in this paper.

II. LITERATURE SURVEY

Literature survey section gives a lot of research works that have been recorded from past few years. This research work presents in the domain of data mining for the most important process for all human beings called as Blood Donation as:

In a paper of PavelBerkhin [7], the web based information system for blood donation and performed extensive research in the field of data mining for organized analysis of the blood bank repositories to get required information, which is important and helpful to the healthcare professionals for a better management of the blood bank facility.

ArunK.Pujari et al. [8] in his work presented how data mining task can be useful to improve the performance of blood donation information analysis using Data mining technique. In this paper, the method of improved k-means clustering is adopted that improves the performance for determining the blood donors information based on the required characteristics as blood group and location at any time when needed.

In June 2009 [9], work is carried out for management information system that helps managers of any organization in providing decision making. This system is about the process of collecting, processing, storing and transmitting the relevant information that is similar to any organization that working as Blood Bank and uses a data mining approach for mine required information.

The author Abdul Nazeer K A. et.al [10] in his research work has given an efficient method for assigning the data-points to the clustersthat reduces the distance calculation. In the original k-means algorithm that computes the distances between the data points from all the centroids for all iteration, that makes the algorithm much expensive for computation performed here. So to reduce the complexity another approach by Fahim uses of two distance functions for this purpose. One of which is similar to the k-means algorithm and another one was based on a heuristics approaches to reduce the number of distance calculations. This algorithm determines the initial centroids randomly as the original one, thus it increasing the accuracy than previous one to determine the clusters.

In a work of Abdur Rashid Khan [11] give one of the improvement of the IT field and the internet that is that is using the information for blood donation, which are available on Web i.e. online and that is usefulfor interconnecting all the blood donor societies in a country using LAN Technology.

An article specifies in [12] gives the information about the data mining techniques that are used on medical and healthcare data. For his work he uses the clinical databases which stores huge amount of information regarding to patients diagnosis, results of laboratory test, patient's prescription, etc. He uses the J48 algorithm and Weka tool for predicting the number of blood donors available between particular age group and blood group.

Chau et al in his paper [13] have analyzed the linkages related to the blood donationbased onlocation and address of the blood donation centers and donors. This research uses donors past donation profiles that help to setup a new blood donation for next for the Hong Kong Red Cross a huge and famous Blood collection center. They provide uniqueness by finding correlations between spatial distance and by finding the incentive for the blood donors. This is specifically helpful for the effective setup of centers with maximal potential of donor-ship.

In the work of Bing [14] extensive analysis, working and implementation of blood bank information systems is performed. Their research can be used as an extensive background of blood bank information systems. The importance of the decision making capability that should be present for effectively running the operations in blood banks is also given by this along with the research that identifies various critical areas of the systems for decision making.

III. DATA MINING AND KDD PROCESS

Data mining is a process for analysis of large amounts of data and taking out the relevant and targeted information. It is also called as the process of extracting large amounts of data [15] as the extraction of useful and comprehensible knowledge, from large data repositories stored in different formats and stored at different places, with the objective of improving the decision for any companies, organizations collecting the information. Data mining is overall process synonymously known as Knowledge Discovery from Databases (KDD). This is especially for the case data mining is applying to the blood bank. The figure 1 of the following shows the process of data mining as the step for the process of KDD.
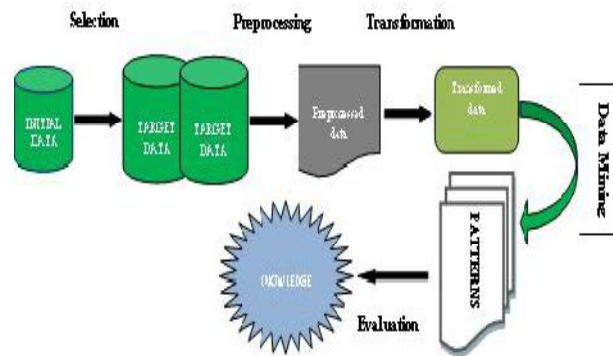
Figure1: Data mining the core of KDD Process [16]

- Selection in case of KDD process includes the data requires for mining process from many different & heterogeneous data sources.
- Preprocessing is the finding incorrect or missing data performing different activities at this time. If there are errors in the data it may be corrected or removed and missing data must be supplied.
- Theprocess of transformation is converting the data into a common format for processing. The use of transformation method like sampling, aggregation, generalizationand Feature selection and attribute subset selection, heuristic method & data, etc.
- Evaluation is the process of most important data mining results are presented to the users. This result is the most important as it shows the usefulness of the evaluated result.

Different kinds of mined knowledge requires different kinds of representation and can be achieved with different data mining process e.g. classification, clustering, association rule,etc.[17].In this paper for our task of blood donation, the discovery of the proper donors which we required for particular times done using the algorithms of classification and clustering process of data mining.

IV. CLASSIFICATION ALGORITHMS

A. Naive Bayes

The Naive Bayes algorithm is the classifier based on Bayes theorem andassumes predictors of the required database information. As it is particularly suited for large amount of dataset contain enormous amount of Inputs our model of data mining for blood donors details is properly suited. Naive bays classifiers algorithm uses all attributes in the dataset and then analyzes individually and make it possibly independent. It is the simple algorithm for classification [19] as compare to other algorithm.

B. J48

J48 decision tree is an implementation of algorithm called as ID3 (Iterative Dichotomiser 3).The Decision tree is mainly the predictive machine learning model form the data and is important in classification technique.J48 algorithm build decision tree on the basis of training dataset. The decision tree uses the fundamental idea that divide the data for easy way of classification. This J48 is useful for many other techniques like error detection and correction of the values, tree pruning. Decision tree is the most useful and one important thing for this is that each attribute in the dataset is used at the time of processing [18]. J48 as like decision tree accept all or both continuous and discrete attributes for classification. This algorithm generates progressively generalization of a decision tree and it gains equilibrium of flexibility and accuracy.

C. CBA (Classification Based Association)

Association rules in the data mining are used to analyses relationships between data from large amount of data. Classification involves learning a function which is capable of mapping some instances of data into distinct classes. Now both the association, classification rule mining can be integrated for forming a framework called as Associative Classification which are referred as Class Association Rules. The use of association rules is restricted to problems where the instances can only belong to a discrete number of classes as association rule mining is mainly possible for attributesbased on some categories. General association rule is not used directly with its attributes. Class association rule is a predictive data mining task. By using the power of the Class Association Rules one can build a classifier [18] for our large datasets.

D. Random Tree

Random tree classification is both work with classification and regression technique problems [19]. For applying the random tree it construct multiple decision tree randomly. In building each tree for decision, this algorithm get remaining feature randomly as in our blood donation process there are multiple attributes. It is a collection of tree predictors called forest as it is not related to decision tree it is randomly predicted.

V. K-MEANS CLUSTERING METHOD

Clustering methods are helpful to discover groups of data records that have similar values or patterns. So these techniques are used in Blood donor's segmentation and in other sectors of business applications [22]. The K-means clustering algorithm is a relatively quick for exploring clusters in datasets. In this algorithm user by setting the number of clusters (k) and each data record is given to the nearest clusters? This procedure is helpful for our blood donor's behavior identification process that can run several times because there are number of clusters and algorithm runs for each cluster one at a time. Therefore, this algorithm can be said as an iterative algorithm [22].

The clustering methods introduced and we can run the k-means clustering algorithm on review dataset with possibly two or more values for number of clusters. The database variables are any parameters based on the data that we are interested to mine using clustering methods. For our Blood donation process these are mainly like age, blood group, previous blood donation time, gender, educational background and marital status, etc. Here, the aim of clustering available dataset is to identify the blood donor behavior and after running k-means clustering algorithm, we can calculate the optimal number of clusters which can be used for our targeted information.

Methodology

Our need of the blood donor and donor's characteristics can be identify by using the clustering algorithm using this steps as follows:

Step 1: Collect data from the blood bank.

Step 2: Apply the K-means Clustering Algorithm for classifying the number of blood donors through the blood group, age, height, gender and location, etc.

Step 3: Extract the Mail ids or phone numbers of the persons from the resultant cluster that satisfying the criteria of blood group and location.

Step 4: Forward the message to the donors as according to the requirement.

In this method, the K-menace clustering algorithm, deals with the multidimensional data attributes such as *di1, di2...dim*, where *m* is the number of attributes or columns in each data value. When data contains the multiple values, we first have to determine the column with maximum range present in the column [20].Then we have to determine the initial centroids [21] from the range of the Column, assuming we have the two-dimensional dataset,
After determining the initial centroids, the data points are divided into k-equal partitions.
Then after that for each iteration, we have to calculate the Euclidean distance between the data point for each centroid, which is just the direct distance between any two points that can be easily measured, from both in 2-D and 3-D space and is the default distance metric which is used in the k-means clustering algorithm [23].

This K-means Clustering Algorithm is stated as with the following steps:

Input: D = {d1, d2,......, dn}
K

Output: Ais a set of *k* clusters.

Steps:
1. Calculate the initial centroids using formula and set the cluster for that centroid.

2. Repeat
2.1 Initially assign each data point for that cluster
2.2 Update the centroid value by calculating the mean of that cluster
This can be done until all data points are assigned to any one of the clusters.

3. Repeat
3.1 Assign each data item *di* to the cluster having the closest centroid;
3.2 Calculate new mean of each cluster meeting convergence criterion.

Here algorithm gives best possible outcome meeting all criteria assign to that cluster. Also, results of applying the clustering algorithm on large database for getting the person to donate their blood based on some of their characteristics and most of them are personal. In the previous section we utilize the classification methods that predict on blood donors database, for example we can used Naive Bayes algorithm, J48 or decision trees and Random trees to examine the blood donor classification [21]. This classification methoddescribed techniques for Blood Group Donors datasets for development a system that is essential for the timely analysis of huge Blood Group Donors database [22]. And for paying attention for describing the blood donor behavior we utilizes clustering methods to help managers of the needy healthcare for increasing the accuracy of their predictions efficiently and properly.

VI. CONCLUSION

From the above study, we get much important knowledge about how the data mining task improves the most important blood donation process. Blood which is vital important and cannot be manufactured by anyone is going on increasing its demand by all over the countries in the world due to the increase in the number of accidents and surgeries. And the need of blood occursmany times on urgent basisand at that time it's not possible to get proper donor information fastly. So, for gaining this information easily and efficiently we can use the data mining technique

that are proper for this task that is useful to provide required information from large datasets. In this paper, we proposed to use Classification technique by using the algorithms like Naive Bayes, J48, Random tree, etc. and classifies the donor's information from the large datasets in classes. After that we use the K-means Clustering algorithm for further dividing the more proper information to make sub-classes of the class that we get form previous Classification algorithm and makes the small groups of information. This clustered form of data gives us the group of donors, which help us to gain proper information about them that will be useful at any time when urgent need of blood will occur.

REFERENCES

[1] Rahman MS, Akter KH, Hossain SH, Basak A, Ahmed SI. Smart blood query: a novel mobile phone based privacy-aware blood donor recruitment and management system for developing regions. IEEE workshops of international conference on advanced information networking and applications (WAINA) 2011. 544-48: 22-25 March 2011.

[2] Li BN, Dong MC, Chao S. On decision making support in blood bank information systems. *Expert Systems with Applications Journal*; 34: 1522–32,2008.

[3] Li B.N, Dong M.C. Banking on Blood. IET Computing & Control Engineering, August/September 2006.

[4] Ashoori M, NajiMoghaddam V, Alizadeh S, Safi M. Classification and clustering algorithm application for prediction of tablet numbers: case study diabetes Disease. *Health Information Management*1392; 10(5): 739-749. [In Persian] ,View 2 Jan 2015.

[5] Ramachandran P, Girija N, Bhuvaneswari T. Classifying blood donors using data mining techniques. *IJCSET*; 1 (1): 10-13, 2011.

[6] K. RAJENDRA PRASAD and Dr. P.GOVINDA RAJULU, "A survey on clustering technique for datasets using
efficient graph structutes" International Journal of Engineering Science and Technology Vol. 2 (7), 2707-2714, 2010

[7] Wen-ChanLee and Bor-Wen Cheng; 'An Intelligent system for improving performance of blood donation',
Journal of Quality, Vol. 18, Issue No. II, 2011.

[8] Arun.K.Pujari, "Data Mining Techniques", Universities Press, (2001).

[9] G. Satyanaryana Reddy et al; 'Management Information System to help Managers for providing decision making in any organization'June 2009.

[10] Abdul Nazeer K A, Sebastian M P, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm," Proceedings of the International Conference on Data Mining and Knowledge Engineering, London, UK, 2009.

[11] Abdur Rashid Khan et and al. 'Web based Information System for Blood Donation', *International Jounal of digital content Technology and its applications*, Vol. 3, Issue No.2, July 2009.

[12] Alaahamouda et al; 'Automated Red Blood Cells counting', *International Journal of Computing Science,* Vol.1, No.2, February 2012.

[13] Michael Chau, Eddie Cheng and Chi Wai Chan, "Data Analysis for Healthcare: A Case Study in Blood Donation Center Analysis", Proceedings of Sixteenth Americas Conference on Information Systems (AMICS), 2010.

[14] Bing Nan Li, Ming Chui Dong and Sam Chao, "On decision making support in blood bank information systems", Expert Systems

[15] Jaiwei Han and MichelineKamber, 'Data Mining Concepts and Techniques', *Second Edition, Morgan Kaufmann Publishers.*

[16] Arvind Sharma, P.C. Gupta, "Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool", *International Journal of Communication and Computer Technologies* Volume 01 – No.6, Issue: 02 September 2012.

[17]Sunita B Aher, LOBO L.M.R.J., *International Conference on Emerging Technology Trends (ICETT)-2011* Proceedings published by International Journal of Computer Applications (IJCA).

[18] S.Asha Rani, Dr.S.Hari Ganesh, "A comparative study of classification algorithm on blood transfusion", *International Journal of Advancements in Research & Technology (IJOART),* Volume 3, Issue 6, ISSN 2278-7763; June-2014 .

[19] Ritika,Aman Paul, "Prediction of Blood Donors Population using Data Mining Classification Technique", *Volume 4, Issue 6,*June 2014.

[20] Pang-Ning Tan, Michael Steinback and Vipin Kumar, *Introduction to Data Mining*, Pearson Education, Transactions on Information Theory, 28(2): 129-136; 2007.

[21] Yuan F, Meng Z. H, Zhang H. X and Dong C. R, "A New Algorithm to Get the Initial Centroids," *Proc. of the 3rd International Conference on Machine Learning and Cybernetics*, pages 26–29, August 2004.

[22] Abdul Nazeer K A, Sebastian M P, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm", *Proceedings of the International Conference on Data Mining and Knowledge Engineering*, London, UK, 2009.

[23] Maryam Ashoori, Zahra Taheri, "UsingClustering Methods for Identifying Blood Donors Behavior", *ICEEE* 2013.

[24] MadhavErraguntla, Peter Tomasulo, "Data Mining to Improve Safety of Blood Donation Process", 47th Hawaii International Conference on SystemScience,   2014.