Study to Make an Application to Improve Blood Donation Process using Data Mining Techniques

Neha Sinha[1] Shradha Maurya[2] Prof. Vaishali Bodade[3]
[1,2]Student [3]Assistant Professor
[1,2,3]Department of Computer Engineering
[1,2,3]Bharati Vidyapeeth College of Engineering, Belpada, Navi Mumbai

Abstract— Blood is the vital thing for human being because there is no alternative for replacing it. At the times of urgency, the availability of the blood creates problem due to which it is necessary for the presence of a system for providing the blood donors to the recipients by collecting the information about the blood donor, health status and blood group predicting continuous behaviour of healthy blood donors in blood transfusion organization. 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 necessary to analyse the blood donor database, along with his information. For retrieving the proper information from large database related to the blood donor's data mining is used to analyse their availability, number of donors and all related information. There are various different techniques and algorithms present in data mining like classification, clustering, association, etc. which suits better for desired task.

Keywords: Blood Donation, Data Mining

I. INTRODUCTION

Nowadays information and computer technology has gaining more importance in medicine and healthcare sectors, as it is needed to make it efficient by computer technology(CT) to database system. As the medical information and healthcare sectors repositories data is complex in the computer technology it explores the use of data mining field.

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 growing in same amount. The blood and their donors are very much important as it cannot be manufactured and only can come from generous donors.

In this study, will study how to give proper and timely delivery of blood by using data mining. The motivation was to create an application to create a common platform between donors, recipient, and blood banks too.

In this study, there are lots of factors to be considered while taking the blood from any donor. As its registration with all his personal information, last donation date and any reaction occurred, blood group etc. These all stored information is possible to access with the different techniques of data mining like classification, clustering, prediction etc. these techniques give the segments of potential blood donors in terms of identifiable characteristics, behaviour patterns, and it also suggest properly at the time of blood requirement in urgent cases.

Therefore, we use data mining tools and techniques for finding best donor which can help the person in need.

II. REVIEW OF LITERATURE

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:

A. Pavel Berkhin:

In a paper of Pavel Berkhin, [1] 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.

B. Arvind Sharma:

In his research work predict the number of blood donor through their age and blood group. In this paper [2], real-time dataset is implemented and measure the accuracy of blood donor prediction using classification algorithm. WEKA tool used for this prediction.

C. Arun K.Pujari:

Arun K. Pujari in his work [3] 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.

D. Michael Chau:

Michael Chau in his paper have analyzed the linkages related to the blood donation based on location and address of the blood donation centers and donors. This research [4] 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.

III. PROPOSED SYSTEM

Our proposed system brings voluntary blood donors and those in need of blood as well as various blood banks with different databases on to a common platform. In this system the donor is able to register themselves, where they have to fill the various information about themselves like blood group, height, weight, last donated at, address and contact

information. Where the person who needs blood will register themselves at first place, and then will search the blood donor in find the donor option, where after using various data mining techniques we will provide the best donor for that blood group with contact information and if we are not able to find the donor, one can even find the nearby blood bank where that blood group is available. Which will increase the efficiency of the recipients to receive the blood.

A. Advantages of Proposed System:

1) Less Time consuming.
2) More efficient than the existing system.
3) The system is flexible, and can work on any android devices.
4) Efficient final result.
5) There are many option for the recipient to look upon which will decrease the problem of unavailability.
6) It is easy and comfortable in use.
7) No training is required.
8) The project is also developed in such a way that the user, new to the system will just have to install the set up and it is ready to go.

B. Methodology:

1) Donor gives information regarding his/her blood group and address and contact number.
2) Hospital look for donors by using their blood group as a key to search from entire donors list, this can be done using Data mining.
3) Steps:

Step 1: Donors give information about them.
Step 2: Apply the k-means clustering algorithm for classifying the number of blood donors through the blood group.
Step 3: Extract the phone number of the persons from the resultant cluster that satisfies the criteria of blood group and donation.
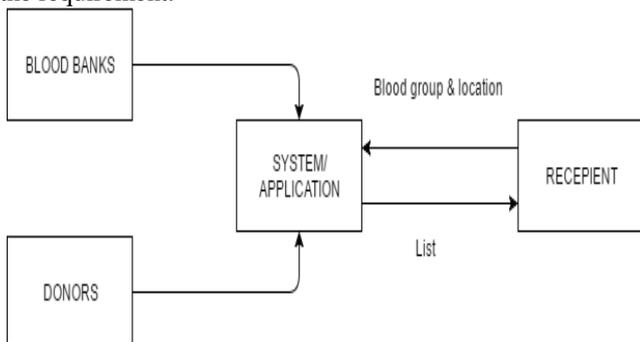Step 4: Forward the message to the donors as according to the requirement.

Fig. 1: Proposed System

IV. ALGORITHMS USED

A. Classification Algorithms:

1) Naive Bayes: The Naive Bayes algorithm is the classifier based on Bayes theorem and assumes 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 as compare to other algorithm.

2) 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. 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.

3) 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 attributes based 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 for our large datasets.

4) Random Tree: Random tree classification is both work with classification and regression technique problems. 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.

B. 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. 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. 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.

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 [5].Then we have to determine the initial centroids [6] 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 [7]. This K-means Clustering Algorithm is stated as with the following steps:

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

K

Output: A is a set of k clusters.

Steps:
1) Calculate the initial centroids using formula and set the cluster for that centroid.
2) Repeat
   – Initially assign each data point for that cluster
   – 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
   – Assign each data item di to the cluster having the closest centroid;
   – 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.

V. 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 occurs many times on urgent basis and at that time it's not possible to get proper donor information fast. 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] Wen-ChanLee and Bor-Wen Cheng; 'An Intelligent system for improving performance of blood donation', Journal of Quality, Vol. 18, Issue No. II, 2011.

[2] Arvind Sharma and Gupta.P.G, "Predicting the number of blood donors through their age and blood group by using data mining tool", International journal of communication and computer technologies (IJCOTS), vol-1, and issue: 02, September 2012.

[3] Arun K. Pujari (2001): Data Mining Techniques, Universities Press.

[4] 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.

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

[6] 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.

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