

Machine Learning Clustering Method for Analysis of Blood Donor Deferral

Shashikala B.M.

Department of Computer
Applications, SJCE,
JSS Science and Technology
University, Mysuru, Karnataka, India

Pushpalatha M.P.

Department of Computer Science
and Engineering, SJCE,
JSS Science and Technology
University, Mysuru, Karnataka, India

Vijaya B.

Department of Pathology, JSS
Medical College,
JSS Academy of Higher Education
and Research, Mysuru, Karnataka,
India

ABSTRACT

The objective of this paper is to classify the deferred donors based on the risk factors. This paper discusses the implementation of clustering technique with related to risk factors associated with the donors for becoming deferred donors. The data for this implementation is collected from local hospitals. The developed system is an unsupervised learning technique. The K- means clustering analysis work is utilized to arrange the blood contributor's depending on the deferral reason. Elbow method used to identify the optimal number of clusters.

General Terms

Algorithms, Design, Experimentation.

Keywords

Deferred donors, Risk factors, Clustering technique, Elbow method

1. INTRODUCTION

Human blood is very important in hospitals to improve the health of people. Hospitals always face the shortage of blood during emergency requirements. Blood bank sectors facing challenges to meet the demand of blood across the country. The demand and supply of blood requirement can be accomplished with increase in voluntary donors. The blood and their benefactors are impressively significant on the grounds that it can't be made and just originate from liberal givers. Voluntary and non-remunerated donors have been considered as safest donors in blood bank sectors [1]. Hence it is important to explore the information of donor's data and convert them to valuable information to retain the voluntary donors for blood donation.

Large volume of data is available in the health sectors, the important data is hidden in the database and the knowledge of extracting the information is necessary. As the volume of the data in the electronic health record increases the analysis of data and converting them into meaningful information is the major requirement in health sector [2]. Data mining tools are used to extract the hidden information for knowledge discovery [3]. To fetch the relevant information from large amounts of data and to analyze the information data mining techniques plays an important role. Extraction of knowledge requires different data mining methods such as association rules, classification, clustering and many others. Extraction of the stored data efficiently is necessary to meet the demand of blood in the blood bank sector [4]. In healthcare and medicine, it is a big challenge to apply data mining algorithms for the extraction of knowledge in stored data [5]. An online blood bank system designed for automatic call routing facility

for the immediate fulfillment of the blood requirement in emergency using donor algorithm. The system proves successful and efficiency in terms of accessibility, speed and security [6].

The aim of study is to search the voluntary deferred donors feature and call them to donate blood in the blood bank during the emergency requirement. In this study clustering techniques are used in order to maintain the blood contributor's information from the dataset so deferred donors return for donation and the elbow method for identifying the optimal value of K in K-means clustering for the dataset.

The organization of the paper is mentioned here: Firstly, the Literature review is described. Methodology section explains about the data set collection, data preprocessing, data analysis, Machine learning algorithm used for study, result analysis and conclusion at the end.

2. LITERATURE REVIEW

Several researchers used various data mining techniques extensively for classification, clustering, association and regression in blood bank management systems. As the stored data increases in the blood bank sector the efficient method to retrieve the data is necessary for the analysis. To analyze the stored data in blood bank information system, data mining techniques plays a very important role for the extraction of data benefitting the blood bank [7].

In order to meet the demand of the blood supply a web-based blood bank management system developed using an algorithm to find the best donor. The system stores the information of the donors and an interface created to bridge the gap between donors and receivers. The developed system benefits both the donor and the receiver [8]. To increase the voluntary blood donation frequency a system has been developed using classification and clustering algorithms, to understand the behavior and intention towards blood donation. Clustering techniques applied to create groups and accuracy of 0.783 is obtained [9]. A system was developed to analyze huge blood group donors' data set. Data mining classifier decision tree used on blood group donors' data set for the timely analysis of donors [10].

To study the demographic factors that influence the blood donation an intelligent modelling technique are used to study the behavior and intention of donors towards donation. Classification ability of algorithms compared with multi-layer perceptron neural network [MLP] and probabilistic neural network [PNN] [11]. To analyze the blood donor information clustering technique with improved k-means used. An improved K-means algorithm produces results with accurate clusters in less computation time to find the donor's

information [12].

A weight-based k-means algorithm proposed to identify the various diseases from the hemogram blood test samples data set. Results of Fuzzy C-means and K-means clustering compared with the results of weight-based k-means algorithm. A weight-based k-means algorithm performs better in terms of cluster accuracy, error rate and execution time than Fuzzy C-means and K-means clustering algorithms [13]. The cost of software development and maintenance can be reduced by using clustering techniques during requirement engineering process. Performance of K-means and Fuzzy C means algorithm evaluated, fuzzy C means proved higher performance compared to K-means [14]. To improve the accuracy of manager prediction in describing the blood donor behavior K-means and two step clustering algorithms are used. Two step algorithm confirms the results of Dunn's index for optimal number of clusters is 2. The two-cluster result describe that most donors are O positive blood group and are men [15].

3. METHODOLOGY

3.1 Data set collection

The data consists of whole blood donation data collected from local hospitals between 2019-2020. The data gathered consists of 8508 instances.

3.2 Data Preprocessing

First the dataset is classified into permanent and temporary deferrals. According to the aim of study temporary deferral instances from the data are selected. The deferral sample

consists of 1003 instances with 11 features. The features are donor number, name, phone number, weight, date, blood group, marital status, age, gender, donor type, risk factors

3.3 Data analysis

The four features age, gender, donor type and risk factors relevant for study are considered. Many people come for whole blood donation at the hospital. During donation some of the donors are not able to donate because of health and others reasons not allow them for donation. As risk factor are associated for not becoming a donor, for the study the risk factor for not becoming donors and return without blood donation are considered. In this the risk factors who are temporarily deferred for donation are considered for analysis. Figure 1 gives the system design of analysis.

The system has been developed with the following datasets obtained from the hospital.

Description of the various features of the data given in Table 1.

Table 1: Description of data

Gender	Male/Female
Age	Age of donors in years
Donor Type	Voluntary/Replacement donor
Risk factors	Alcohol, anaemia, hypertension, medication, underweight, surgery, skin allergy, tattoo

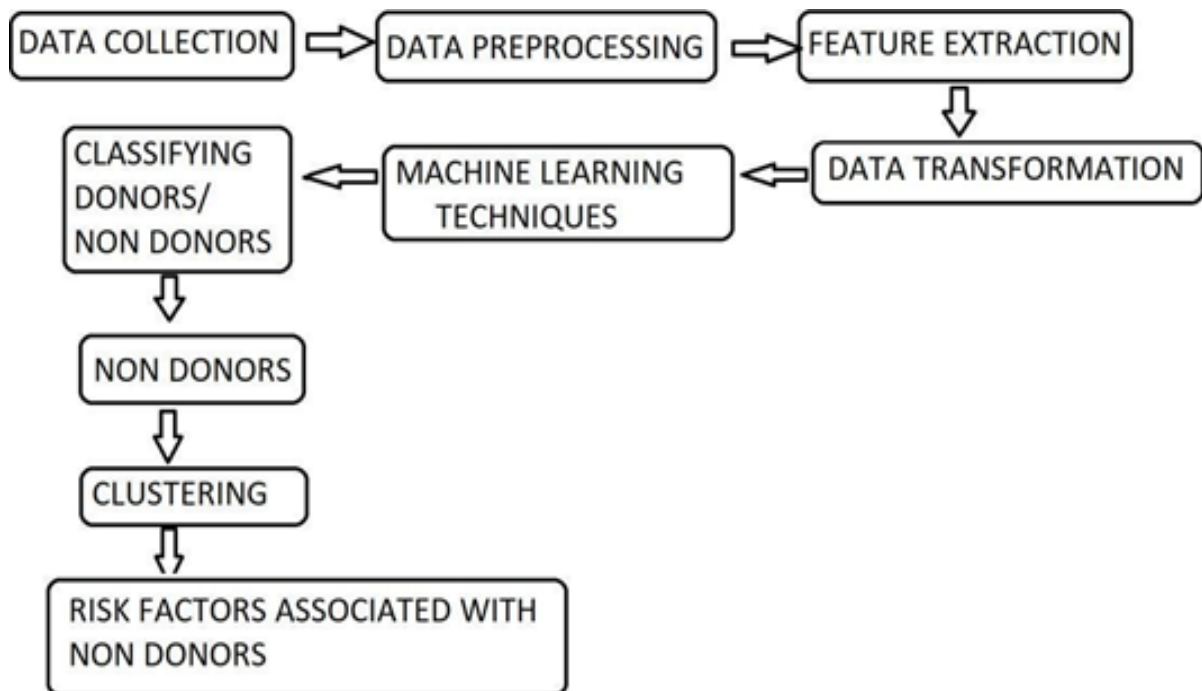


Fig 1: System Design

3.4 Machine Learning Algorithm used for study

K-means clustering is an unsupervised machine learning algorithm. Algorithm forms groups of the objects which are similar. The clustering algorithm used to cluster the input data into classes. Elbow method used to find the value of k for k-

means clustering.

4. RESULT ANALYSIS

The k-means clustering algorithm for the dataset with k=2,3,4,5,6 is used. The data set attributes are age, reasons, donor type and gender. The aim of the experiment is identifying the reasons for deferral in blood donation using

clustering.

To find the optimal number of clusters the elbow method is used. The elbow method selects the optimal number of clusters for the dataset. The elbow method perform K means clustering on the dataset for range of values for each of the k values, average distances to the centroid for all data points calculated. Average distance from the centroid falls suddenly for k=2(i.e., the sum of squared distances). So, the optimal number of clusters formed should be 2 for the dataset. Fig 2 gives the elbow method used for selecting number of clusters.

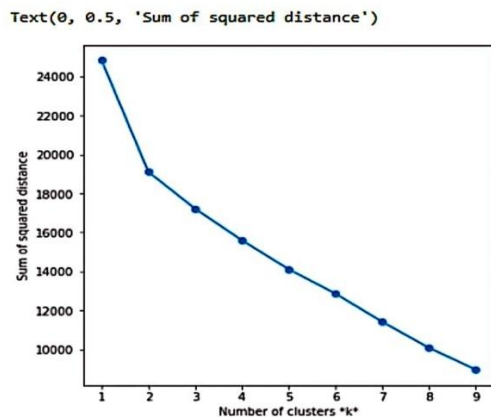


Fig 2. Elbow method for selecting k value for cluster

The clustering results show that the study dataset is segmented into two clusters. The aim is to identify the clusters based on the risk factor in each donor, to distinguish deferred donors with risk factors from other clusters. Cluster 1 contains 351 samples of whole blood donors where 23.9687 value for average age, all of them are voluntary donors and men. And most of them are deferred with reasons of alcohol (35.18%). Cluster 2 contains 652 samples of whole blood donors where 29.483 value for average age, all of them are voluntary donors and men. And most of them are deferred with reasons of anaemia (28.18%).

The two-cluster analysis has shown men tend to donate more and all are voluntary donors. Also, the most deferred reason for donation is anaemia. Clustering methods help the managers to identify the risk factors and call back the donors based on the deferred reasons. The results of cluster help the blood transfusion services to call back the deferred people once they become eligible for donation. For example, the deferred donors for alcohol reason can be called back after 48 hours during the emergency requirement of blood based on blood group. From this losing voluntary non-remunerated donors can be avoided. People are more tend to donate the blood but because of deferred reasons they are unable to donate. The results of cluster help the blood transfusion services to identify the risk factors for the deferral for donation, this module can be used to keep the track of the deferred donors.

5. CONCLUSIONS

The result of cluster analysis has shown satisfactory results with sensitivity and specificity analysis. The idea of using unsupervised clustering analysis is to identify the deferred reasons for donation of blood. Cluster results are evaluated with the test set data which has shown optimal result. The results of cluster can be used by blood bank management people when there is a requirement of blood supply. Analysis

of two cluster shows the most of the donors are men. The analysis helps the health workers to identify health issues in population and people with anaemia can be suggested for intake of iron. The work can be further extended to identify various reasons for deferral in blood donation by considering the datasets of various places.

6. REFERENCES

- [1] Rahman, M.S., Akter, K.A., Hossain, S., Basak, A., and Ahmed, S.I. 2011. "Smart blood query: a novel mobile phone based privacy aware blood donor recruitment and management system for developing regions", IEEE Workshops of International Conference.
- [2] Naji Moghadam, V., Ashoori, M., Alizadeh, S., and Safi, M. 2012. "The Classification Algorithm for number of tablet usage prediction: case study diabetes", The sixth Iran Data Mining Conference, Tehran: Iran Data Mining.
- [3] Rashid Mehrabadi, E. and Pedram, M.M. 2010. "Blood Donors Classification and Identifying Future Donors", The Fourth Iran Data Mining Conference, Sharif University of Technology, Tehran, Iran.
- [4] Sobia Zahra, Mustansar Ali Ghazanfar, Asra Khalid, Muhammad Awais Azam, Usman Naem b, and Adam Prugel-Bennett. 2015. "Novel Centroid Selection Approaches for Means -Clustering Based Recommender Systems", Elsevier.
- [5] Rajput, A., R.P. Chandel, N., Solanki, D.D., and Soni, R.. 2009. "Approaches of classification to policy of analysis of medical data", International Journal of computer science and network security.
- [6] K Selvamani. And Ashok Kumar Pai. 2015. "A novel technique for online blood bank management", International Conference on Intelligent Computing, Communication & Convergence, Elsevier.
- [7] Bhardwaj, A., Sharma, A., and Shrivastava, V.K. 2012. "Data Mining Techniques and Their Implementation in Blood Bank Sector –A Review", International Journal of Engineering Research and Applications (IJERA), Vol. 2, no.4, pp. 1303-1309.
- [8] Shashikala, B M., Puspalatha, M P., and Vijaya, B. 2017. "Web Based Blood Donation Management System (BDMS) and Notifications", International conference on Cognitive Computing and Information Processing, CCIP (Springer).
- [9] Chen Lee, W. and Cheng, B. W. 2011. "An intelligent system for improving performance of blood donation", Journal of Quality, vol.18, no.2, pp.173-178.
- [10] Ramachandran, P., Girija, N., and Bhuvaneshwari, T. 2011. "Classifying Blood Donors Using Data Mining Techniques", International Journal of Computer Science & Engineering Technology (IJCSSET), vol. 1, no.1, pp.10-13.
- [11] Mohamed, M. Mostafa. 2009. "Profiling blood donors in Egypt: A neural network Analysis", Expert Systems with Applications an International Journal, vol. 36, pp. 5031-5038.
- [12] Bondu Venkateswarlu, G. S., V Prasad Raju. 2013. "Mine Blood Donors Information through Improved K-Means Clustering", International Journal of Computational Science and Information Technology.

- [13] Vijayarani, S., and Sudha, S.2015. “An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples”, *Indian Journal of Science and Technology*.
- [14] Sheshasaayee A., and Sharmila, P.2014.” Comparative study of fuzzy C means and K means Algorithm for requirements clustering”, *Indian Journal of Science and Technology*, vol.7, no.6, pp.853-857
- [15] Maryam Ashoori, and Zahra Taheri.2013 “Using Clustering Methods for Identifying Blood Donors Behavior”,5th Iranian conference on electrical and electronics engineering, IEEE.