Real-Time Blood Donor Management Using Dashboards Based on Data Mining Models

Shyam Sundaram[1] and Santhanam T[2]

**[1] PG & Research, Dept. of Computer Science, DG Vaishnav College
Chennai 600106, Tamil Nadu,India**

**[2] PG & Research, Dept. of Computer Science, DG Vaishnav College
Chennai 600106, Tamil Nadu,India**

Abstract

This study uses data mining modeling techniques to examine the blood donor classification and extending this to facilitate the development of realtime blood donor management using dashboards with blood profile and geo-location data. This enables decision makers the ability to manage and plan the blood donation activities based on key metrics. This capability provides the ability to plan effective targeted blood donation campaigns. The scoring algorithm implemented for the dashboard also helps in the optimized deployment of budget resources and budget allocation determination for blood donation campaigns.

Keywords: **Dashboards, Blood bank, blood transfusion, blood donor, data mining, classification algorithms, Decision support, blood bank information systems**

1. Introduction

It is essential for healthcare systems to have a constant balance of supply and demand for blood products. These play a critical role in the saving and the extension of life. Dashboards are of critical importance to help achieve these objectives. There are two critical aspects of these dashboards that will be covered in this paper. Firstly the ability to have realtime data that identifies the donor profiles based on their patterns of donorship. Secondly rolling up the profile to demographic level both summary and detail level. In specific the linkage to geo-location will be demonstrated. This paper dwells into the effective dashboard creation using data mining techniques coupled with geo-locational linkage. The objective of this research is to effectively translate data mining derived models to enable decision making with the help of dashboards.

This paper is organized as follows. Section two deals with the introduction to blood donorship and relevant peer research in the context of this paper. Section three describes details with reference to the dataset and analysis. The conclusion is given in the final section.

2. Blood Donorship

Major headings are to be column centered in a bold font without underline. A donation is when a donor gives blood for storage at a blood bank for transfusion to an unknown recipient. These can occur at a number of locations including blood donation centers,mobile camps, mobile vans,etc. There a number of types of blood donations such as voluntary blood donation programme. This is the foundation for safe and quality blood transfusion service as the blood collection from voluntary non-remunerated blood donors is considered to be the safest. In order to augment voluntary blood donation in developing countries like India[1] is based on well defined frameworks and operational guide for organizations for this important activity. International healthcare research bodies have extensive frameworks that address context of blood management[2]. In developed countries there are dedicated organizations that have effective blood donor management processes. One such example is the U.S. department of defense (DOD) , which uses an enterprise blood Management software that will manage the blood supply chain including donor management, blood collections, testing, distribution and transfusion. Additionally this also provides a proactive delivery of information and services through a web portal[3].

2.1 Relevant Peer Research

Santhanam et al[4][5] extended the nominal definition based on a standard dataset to derive a CART[6] based decision tree model based on standard donorship. This analysis helped identify the attributes that classify a regular voluntary donor (RVD) in the context of a standard dataset. This provided an extended RVD definition based on the donor definition (along with the application of CART) provides a standard model to determine the donor behavior and provides the capability to build a classification model. This additional nominal class can be easily computed based on the statistical definitions and help assist in decision making.

Chau et al[7] have extensively analyzed the linkages related to the blood donation to the location of the blood donation centers.This research was carried out using donor's past donation profiles to help setup a new blood donation center for the Hong Kong Red Cross.Their findings provide correlations between spatial distance and the incentive for the blood donors which is the uniqueness of this research. This specifically helps in the effective setup of centers with maximal donorship potential.

Saberton et al[8] have extensively analyzed the linkages related to the blood donation to the location of the blood donation centers.Their findings provide correlations between spatial distance and the incentive for the blood donors. This specifically helps in the effective setup of centers with maximal donorship potential.

Bing et al[9] have extensively analyzed the working and implementation of blood bank information systems. Their research provides an extensive background of blood bank information systems. The research also talks about the importance of the decision making capability that is required for effectively running the operations in blood banks. The research also identifies various critical areas that are required for the systems to also have in order to enable decision making.

3. Analysis

3.1 About the Dataset

The blood transfusion dataset (taken from the UCI ML repository)[10] is based on donor database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan. The center passes their blood transfusion service bus to one university in Hsin-Chu City to gather blood donated about every three months. This dataset is derived from I-Cheng Yeh[11].

The data set consists of 748 donors at random from the donor database. These 748 donor data, each one included R (Recency - months since last donation), F (Frequency - total number of donation), M (Monetary - total blood donated in c.c.), T (Time - months since first donation), and a binary variable representing whether he/she donated blood in March 2007 (1 stand for donating blood; 0 stands for not donating blood). There is an imbalance in that the people who have donated blood in 2007 accounts for only 24% in the dataset.

This dataset has been extended to accommodate the following attributes. RVD a boolean attribute that is computed based on the original attributes along with definitions[1]. Additionally a geo-location information was added in the syntax of latitude:longitude. This was randomly assigned to locations in India for analysis. Please note the data used is to be considered only for demonstrative purposes.

3.2 Analysis

The past research of [5] resulted in creating an extended RVD definition based on the donor definition (along with the application of CART) provides a standard model to determine the donor behavior and provides the capability to build a classification model. The ability to easily compute this based on statistical and definition data provided by frameworks[1]. additional nominal class can be easily computed based on the definition. The results of the decision tree help refractor the definitions of the RVD with the following offsets. These have had been defined using suggestive definitions[5]. The dataset is now corrected to these offsets and used for this analysis.

```
IF ( (Frequency > 18.5 ( times)  AND
Recency < 8.5 (months)  )
        RVD = TRUE
ELSE
        RVD = FALSE
```

This results in a finer refinement to the RVD model. The RVD confusion matrix(using the tool weka[12]) post this is provided in the following table.

Table 1: Revised RVD confusion matrix

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Class 0 (not RVD)	0.92	0	1	0.92	0.96	0.96
Class 1 (RVD)	1	0.08	1	1	1	0.96
Weighted Avg.	1	0.08	1	1	1	0.96

Table 2: Previous RVD confusion matrix

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Class 0 (not RVD)	1	0.06	0.99	0.92	0.99	0.97
Class 1 (RVD)	1	0.08	1	1	1	0.96
Weighted Average	1	0.08	1	1	1	0.96

The comparison between the RVD before the offset and with it indicate an overall stability to the model with delta change to a better true positive rate for a non RVD and also small increase in the FP rate the non RVD.The inclusion of geospatial location along with the donor data provides critical indicative identification of the RVD. This allows the capability to search by geo-locational attributes which enables targeted blood donation program management including aspects related to logistics and infrastructure linkages. Additional linking to census and demographic information[8] allows the effective determination of blood donor profiles with capabilities to drill-down to the appropriate levels.

Please note the this analysis has been developed using random geo-locational values (dummy values) which have helped to provide a meaningful endpoint of this analysis.

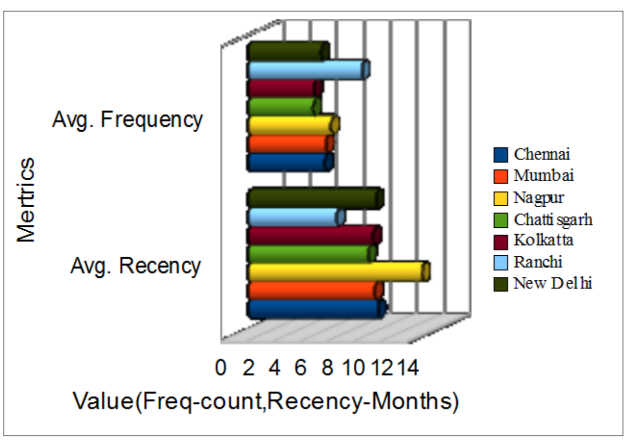

Fig. 1 RVD Geo-locational profiles.

This is further analyzed by a perspective at the overall dashboard across the the indicators and ranking the locations by scores. The algorithm for the dashboard is indicated as follows. Figure 2 provides the implementation of this algorithm.

Geo-location RVD Scoring Algorithm
Step 1: Loop through each unique location L (latitude,longitude) based on geographic division (such as state, district and city).
Step 2: For each location L compute the average frequency,average recency and total RVD count.
Step 3: Calculation of Location level summary scores for the recency,frequency and RVD across the locations.
The RecencyScore (location) is computed as the Rank in descending.
The FrequencyScore (location) is computed as the Rank in ascending.
The RVDScore (location) is computed as the Rank in ascending.
Step 4: Plot the this score in the chart with scores on the X – Axis and locations on the Y-axis.

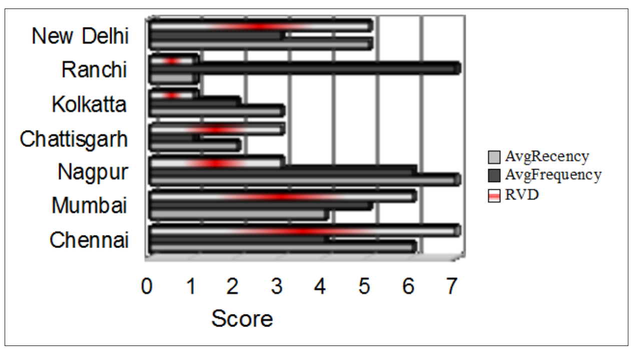

Fig. 2 Location-wise indicator patterns.

The results in comparison with the earlier model [5] reveal an improvement in the true positive rate for RVD class along with a delta increase in the false positive rate.

4. Conclusions

The dashboard indicated in figure 2 provide a quick and relevant score of the the geographic locations based on the RVD profile key indicators. The geo-location RVD scoring algorithm can be modified to rollup to additional attributes as well as handle the requisite geographic division strategy. This capability enables this to be linked effectively to the census tracts as well as health profile systems that enable drill-down to finite levels of information for effective blood donor management. This paper provides a complimentary capability to the recent research[7][8][9]. The application of this across a larger dataset and linkage to both demographic and census tracts will enable the ability to identify meaningful patterns of blood donorship that will assist in the better management.

This provide critical decision makers the ability to make planned decisions. This demonstrates a viable mechanism to manage blood donorship. In specific this helps address the optimized deployment of budget resources related to blood donations drives. This also assists policy makers plan the required budget allocation for overall blood donation related activities in a addressing targeted goals. Such techniques assist in the decision support for healthcare organizations.

Future work will be focused on further enhancing these models to allow integration with blood donor management systems including innovative ways of visualization. The current implementation of the RVD model can also be implemented with other relevant attributes. Similar strategies can also be adopted for other healthcare domains.

References

[1] Ministry of Health and Family Welfare,National AIDS Control Organization Government of India. Name, Voluntary Blood Donation Programme , 2007.

[2] MedlinePlus, U.S.National Library of Medicine, NIH,USA.Blood Transfusion and Donation ,http://www.nlm.nih.gov/medlineplus/bloodtransfusionanddonation.html, 2011.

[3] The Armed Services Blood Program (ASBP, ,http://www.militaryblood.dod.mil/About/default.aspx 2011.

[4] Shyam Sundaram and T. Santhanam, "Classification of Blood Donors using Data Mining", Proceedings of the Semantic E-Business and Enterprise Computing , 2009, pp. 145-147.

[5] Shyam Sundaram and T. Santhanam, "Application of CART Algorithm in Blood Donors Classification", Journal of Computer Science, Vol. 6, No. 5, 2010, pp. 548-552.

[6] L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone. Classification and regression trees, Wadsworth ,1984.

[7] 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.

[8] Saberton, P. J., Paez, A., Newbold, K. B., and Heddle, N. M., "Geographical Variations in the Correlates of Blood Donor Turnout Rates: An Investigation of Canadian Metropolitan Areas", International Journal of Health Geographics ,Vol. 8, No. 56, 2009.

[9] Bing Nan Li, Ming Chui Dong and Sam Chao, "On decision making support in blood bank information systems", Expert Systems with Applications,Vol. 34, No. 2, 2009, pp. 1522-1532.

[10] A. Asuncion and D.J. Newman, UCI repository of machine learning database" www.ics.uci.edu/_mlearn/MLRepository.html,, 2007.

[11] I.C. Yeh , K.J. Yang, Ting and T. Ming, "Knowledge discovery on RFM model using bernoulli sequence", Expert Systems with Applications,Vol. 36, No. 5, 2009, pp. 5866-5871.

[12] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques, Wadsworth ,Morgan Kaufmann, San Francisco, 2005.