6 Real-Time Blood Donor Management using Dashboards based on Data Mining Models

6.1 Introduction

This study uses data mining modelling techniques to examine the blood donor classification and extending this to facilitate the development of real-time 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.

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 real-time 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
6.2 Details

The management and the maintenance of an effective pool of regular blood donors is an important element for protecting public health. The planning and development of blood donation services requires the motivation for repeat donations through the RVD. This is also linked to effective management of the infrastructure and locational attributes as well. These can occur at a number of locations including blood donation centers, mobile camps, mobile vans, etc. There are 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 [MF07] and in Europe [D13] 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 [M13c]. In developed countries there are dedicated organizations that have effective blood donor management processes. One such example is the Armed Services Blood Program (ASBP) [A13b] which plays a key role in providing quality blood products for Service members and their families. This is a joint operation among the military services (Army, Navy, Air Force).
Review of Literature

Michael [ME+10] 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. The research identified the relationship between donor’s demographic information and their activeness in blood donation, the relationship of geographic distribution of blood donation centers and the frequencies of blood donations, and the characteristic of each blood donation center. The findings of the research help identify the best location for a new blood donation center which can increase the usage of the new center. The dataset contains 775,690 records for blood donation. This included the donor identification, their demographic information, information about each blood collection, and the information (about the site making donations). The study found by categorizing all records based on the 18 districts, the average of total blood donation per total number of donors is 2.23 (in the past four years). By restricting the venue of blood donation to donation centers only, this number changes to 2.79. One of the findings of the research was that people in the centers make blood donation at higher frequency compared with the average (19.44% increase in blood donation frequency in the centers when compared to other mobile donation units).

Saberton et al [SP+09] have extensively analyzed the linkages
related to the blood donation to the location of the blood donation centers. The study used geo-coded blood donor and donor clinic data (from Canadian Blood Services for the fiscal year 2006-2007) indicating the total number of donors for each Canadian postal code, excluding the province of Quebec. Potential correlates of blood donation are selected based on social and economic characteristics (as well as descriptors of city size, geographical location in the urban hierarchy, measures of accessibility and capacity of donor clinics). Some of the findings included that as the proportions of working age population and seniors increase, the donor yield tends to decrease. Other variables that correlate positively with donor rates are the proportion of English speakers, highly educated individuals, and the proportion of people employed in health-related occupations. Their findings provide correlations between spatial distance (around 4 km) and the incentive for the blood donors. This specifically helps in the effective setup of centers with maximal donorship potential.

Guo et al [GW+13] studied the long-term return behaviour of Chinese whole blood donors. This study used data from 54,267 Whole Blood (WB) donors (January 1 and March 31 2008) at the five blood centers in China (and were followed for 2.5 years). Logistic regression was conducted to identify factors associated with their return behaviour. In addition a recurrent-event Cox proportional-hazard model was used to evaluate the overall effect of demographic variables and return behaviour among first-time donors. The results Encouraging first-time donors to make multiple donations is important for keeping adequate blood supply. The finding proposed that first-time and repeat donors shared the same
predictors for return. This indicates that retention strategies on repeat donors may be effective on first-time donors.

Pedro et al [PM+13] discuss about methodological guidelines for reducing the complexity of data warehouse development for transactional blood bank systems. These provide approaches from the IT design and implementation perspectives to the context of blood banks. Mpuntsha and Reddy [MR13] provide experience and learning in blood supply management from Africa. Their study is related to the blood supply data systems used to analyse and manage blood collection and distribution in SANBS. The South African National Blood Service (SANBS) supplies approximately 85% of the country’s blood products. The research publication offers some detailed analysis in areas such collection and wastage of blood products. Some key highlights include the establishing a Score Board IT dashboard that provides improvements to ensure daily update and action triggers to ensure speed responses to variations in the supply or needs from the hospitals.

Cork et al [CC+13] in their patent applications discuss about the mobile applications for medical devices. They describe usage scenarios of interfaces for blood collection device operators and blood center administrators. The patent covers details on a mobile device including an interface listing one or more blood collection (or processing) instruments under control of the operator. Poon et al [PL+13] researched variations of motivation between weekday and weekend donors and their association with distance from blood donation centres. This provides insights into critical eco-system related attributes to blood donation. In this study donors giving
blood in January 2012 were invited to participate in a cross-sectional study by completing an anonymous online questionnaire. Residence and work (or school) locations were collected together with demographics and donation histories. The study compared the motivated donors with less motivated ones in terms of their timing of blood donation and the spatial relationship with the donor centres. The results proposed weekday centre donors were less likely to have returned for blood donation within a year and intend to donate in the following 6 months. One of the findings was regardless of the level of donor’s motivation for blood donation, fewer weekday donations were made if the distance between location of school or workplace and donor centre increased. Shaz et al [SH+13] provide insight into the critical requirement of blood management and its linkage to healthcare systems for overall better management.

Bing et al [BM+09] 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. Santhanam et al [SS09b, SS10a] extended the nominal definition based on a standard dataset to derive a CART [BF+84] 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. The research suggested an extended RVD definition based on the donor
definition (along with the application of CART) provides a standard model to determine the donor behaviour 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.

6.3 Experiments

About the Dataset

This research continues to develop on the blood transfusion dataset [BK13, YY+09] which is based on donor database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan. 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 [MF07]. 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.

The past research [SS10b] 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 behaviour and provides the capability to build a classification model. The ability to easily compute this based on statistical and definition data suggested by frameworks [MF07]. 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 (figure 3.2). The dataset is now corrected to these offsets and used for this analysis. Weka [G95] has been used for the analysis of the algorithms.

IF \((Frequency > 18.5\text{(times)} \ AND \ Recency < 8.5\text{(months)})\)

\[RVD = \text{TRUE} \ (6.1)\]

ELSE

\[RVD = \text{FALSE}\]

This results in a finer refinement to the RVD model. The RVD confusion matrix (post this refinement) is detailed in the following table.

**Table 6.1: Revised RVD CART Confusion Matrix**

<table>
<thead>
<tr>
<th>Class</th>
<th>TP</th>
<th>FP</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 0 (not RVD)</td>
<td>0.92</td>
<td>0</td>
<td>1</td>
<td>0.92</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Class 1 (RVD)</td>
<td>1</td>
<td>0.08</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>1</td>
<td>0.08</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
</tr>
</tbody>
</table>

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 geo-spatial location along with the donor data provides critical indicative identification of the RVD. This allows the capability to search by geolocational attributes which enables targeted blood donation program management including aspects related to logistics and infrastructure.
Additional linking to census and demographic information \cite{SP+09} allows the effective determination of blood donor profiles with capabilities to drill-down to the appropriate levels.

Please note this analysis has been developed using random geo-locational values (random dummy approximate Indian geography locations of approximate regions of Chennai, Mumbai, Nagpur, Chattisgarh, Kolkatta, Ranchi and Delhi) which have helped to provide a meaningful illustrative approach of this research.

This is further analyzed by a perspective at the overall dashboard across the indicators and ranking the locations by scores. The algorithm for the dashboard is indicated as follows. The flow chart of the steps also provides the approach required for this computation (figure 6.1). The results of application of this algorithm provide the details (table 6.2) of the recency and frequency in conjunction with the RVD (figure 6.2).

**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 Xaxis and locations on the Y-axis

Some additional key points to note on this algorithm.

• It must be noted that the drill-down criteria outlined in step 1 maybe customized basis the requirement.

• The drill-down criteria allows the aggregation and the computation of the key value indicators. The implementation of the algorithm needs to factor this as a configurational aspect.

• The location in this algorithm is taken as a geo-locational value for precision, but this may also be configured to key references.

• The scores obtained need to be defined through configuration at specified refresh periods for standardizations.
Figure 6.1: Geo-location RVD Scoring Flowchart
Figure 6.2: RVD Geo-locational profiles

Start

Loop through each unique location \( L \) (latitude, longitude) based on geographic division (such as state, district and city).

For each location \( L \) compute the average frequency, average recency and total RVD count.

Calculation of Location level summary scores for the recency, frequency and RVD across the locations.

* Compute for each location
  (a) RecencyScore & Rank in descending.
  (b) FrequencyScore & Rank in ascending.
  (c) RVDScore & Rank in ascending.

Plot the this score in the chart
X-axis - scores
Y-axis - locations

End
Table 6.2: RVD Geo-locational Profiles Details

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>2.42</td>
<td>7.89</td>
<td>1973.68</td>
<td>27.47</td>
<td>0</td>
</tr>
<tr>
<td>L2</td>
<td>3</td>
<td>10.5</td>
<td>2625</td>
<td>28.07</td>
<td>2</td>
</tr>
<tr>
<td>L3</td>
<td>2.79</td>
<td>8.53</td>
<td>2131.58</td>
<td>29.47</td>
<td>1</td>
</tr>
<tr>
<td>L4</td>
<td>2.73</td>
<td>9.27</td>
<td>2316.67</td>
<td>30.6</td>
<td>2</td>
</tr>
<tr>
<td>L5</td>
<td>2.87</td>
<td>7.4</td>
<td>1850</td>
<td>24.87</td>
<td>1</td>
</tr>
<tr>
<td>L6</td>
<td>2.75</td>
<td>9.75</td>
<td>2437.5</td>
<td>28.81</td>
<td>1</td>
</tr>
</tbody>
</table>

Visualization Techniques Used

The results generated by the algorithm are converted into powerful visualizations that provide clarity. The framework used for the visualization consisted of the R-project [R13] and Google Visualization (googleVis) [G13c] library Application Programmer Interface (API). The source code has been illustrated in chapter 6 (figures 9.9 and 9.10 in under section 9.2) for reference. The data used for this code base is the result of the algorithm (table 6.2).

The googleVis library is a R-project based package providing an interface between R and the Google Chart Tools. This research has used the Geo Map (via gvisGeoMap) and the Map (via gvisMap) API’s. The Geo Map API’s allow the data to be overlayed on a map of a country, continent, or region map, with colors and values assigned to specific regions. the Map API’s displays a map using the Google Maps API. The data values are displayed as points on the map. Data values can be coordinates (Lat-Long pairs) or addresses. These visualizations have been produced for illustrative dashboards.
for this research (figures 6.4 to 6.6).

6.4 Results

The results in comparison with the earlier model [SS10a, SS10b] suggest an improvement in the true positive rate (TP) for RVD class along with a delta increase in the false positive rate (FP).

- True Positive (TP) rate of class not RVD has got refined to 0.92 (from a higher value of 1)
- False Positive (FP) rate for class not RVD has a refined value of 0 (from 0.06)
- Overall precision has increased for the class (not RVD)

Figure 6.3: Location-wise Indicator patterns
Figure 6.4: Advanced Geo-locational Dashboards - Sample 1

Figure 6.5: Advanced Geo-locational Dashboards - Sample 2
6.5 Conclusion

The dashboard (figure 6.3) provides a quick and relevant score of the geographic locations based on the RVD profile key indicators.

- The analysis can be represented in more dashboards (figures 6.4 and 6.5) with geo-location representation (using R project[R13] and googleVis API [G13c]). Advanced versions of this dashboard can be customised for mobile based usages (figure 6.6).

- The geo-location RVD scoring algorithm can be modified to roll-up to additional attributes as well as handle the requisite geographic division strategy.

- This capability enables the linking to census tracts as well as health profile systems that enable drill-down to finite levels of information for effective blood donor management.

- In specific this research provides complimentary capability (SIBAS [BM+09]) for advanced reports and dashboards. The ability to manage and plan blood campaign planning are also areas that can be possible looked at.

- The RVD profiles with geo-location capability help in the planning and setup of blood donation centers for optimal requirements. The suggested approaches for geo-location scoring can be applied in extension to the work done by Michael et al [ME+10].
• The effective and optimal spatial distance of blood donation centers to blood donors can be also be further enhanced by the understanding of the RVD along with their geo-locational details. This provides a logical compliment to the work done by Saberton et al [SP+09]. This specifically helps in the effective setup of centers with maximal donorship potential.

• These findings can be applied to larger datasets with linkage to demographic and census tracts which will enable the ability to identify meaningful patterns of blood donorship.

Figure 6.6: Mobile Dashboards (Prototype)
These dashboards demonstrate a viable mechanism to manage blood donorship. In specific this helps address the optimized deployment of budget resources related to blood donations drives. Another benefits is these approaches assists policy makers plan the required budget allocation for overall blood donation related activities to meet the specific goals. Such techniques assist in the decision support for healthcare organizations in the following areas.

- Define, design and setup targeted activities for donor recruitment and retention
- Analyse the donor management process for optimal operations
- Manage the operations and the demand for blood products