7 Blood Donor Campaign Planning and Management using Data Mining

7.1 Introduction

This study uses data mining modelling techniques to facilitate effective blood donor campaign management. In specific the development of key metrics that assist effective blood donation campaigns are considered. These metrics also help provide insights into blood donor behaviour (from data in blood donor management systems). The suggested algorithms provide effective analytics that enable decision making for optimized deployment of budget resources and budget allocation for blood donation campaigns.

Blood centers are challenged with the need for effective strategies for blood donor strategies. Obtaining adequate supplies of blood from donors is one such challenge area. The other challenges include campaign planning, donor retention and maintaining regular donors to meet the optimal levels (within optimal operational cost) is of importance. Blood banks have computer based software (blood donor management systems) that enable the tracking of activities and transactions (through IT enablement) in the center.

The data that is captured needs to be turned into analytics (to enable the review and decision making) is much required to meet these objectives. Taking advantage of the blood donor management data to identify critical patterns that help plan and execute effective blood donor campaigns is critical activity. Alongside with these enabling the overall privacy of the blood donor and transactions
are also critical from an overall requisite compliance as mandated. Blood donation and decision to donate blood is complex hence laying an information driven approach to the management of campaigns is critical and much required.

7.2 Details

Application of data mining (DM) techniques that help in the optimization of the process of screening for donors by helping to identify regular voluntary donor (RVD). Providing a real-time tracking of key attributes and critical indicators helps decision making and provides an effective basis to plan for campaigns.

Ability to enable decision making for blood bank management in specific towards the recruitment drives. The ability to effective decision making through technology enabled dashboards improve the optimal utilization of resources. While there has been extensive blood donor data that is available through blood banks and healthcare data (government and private), blood donor behaviour is complex.

Review of Literature

The earlier chapters also provide valuable references to work done in areas of blood donation and applications of DM. In this we present some very specific references to this chapter. Here we present some critical references to this chapter. Zabihi et al [ZR+11] have employed fuzzy sequential pattern algorithms to the blood transfusion data set. The research provides a relationship between the donating times and the donating in a special month. The
analysis provides rules that are derived that help to predict a donor behaviour in the future. Wen-chen and Bor-wen [WB11] have developed a system employing clustering and classification algorithms to determine the disparities in blood donation behaviour among the present donors. They have built a model to predict their intentions towards donation. The objective is to increase voluntary blood donation frequency. Their research employed clustering analysis. The proposed system helps blood center management a perspective into understanding in blood donor’s intentions and behaviours.

Michael et al [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.

Bing et al [BM+09] have researched into 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 identified various critical areas that are required for the systems to enable decision making. Saberton et al [SP+09] have analyzed the linkages related to the blood donation to the location of the blood donation centers. Their findings provide useful patterns and correlations between spatial distance and the incentive for the blood donors. This provides suggestive pointers for setup of centers with maximal donorship
potential. France et al [FF+07] have employed the theory of planned behaviour coupled with path analysis to predict blood donor intention. The research was conducted on the basis of an online survey across 227 experienced donors. The survey helped analyze blood donor intention with inputs relating to measurements such as attitude, subjective norm, personal normal norm, self efficacy, behavioural intention, blood donations reaction inventory and blood donor satisfaction. Self-efficacy showed the strongest positive relationship to donation intention followed by attitude, subjective norm, satisfaction and personal moral norm.

Santhanam et al [SS10a, SS10b, SS11] extended the nominal definition based on a standard dataset to derive a CART 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 suggested an extended RVD definition based on the donor definition 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.

7.3 Experiments

About the Dataset

This research continues to use the blood transfusion dataset [BK13, YY+09] is based on donor database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan. Further details on this dataset
are detailed in section 10.3. The analysis has been done using the WEKA [G95] with the development of classification models on this dataset and in specific applying the CART (Classification and Regression Trees) classification algorithm [BF+84]. This dataset has been extended to accommodate the following attributes. RVD and geo-locational attributes have extended this data set as described in the past chapters. Please do note these extensions have been done for illustrative purposes. The dataset has also been additionally extended to provide the following. These will be described in detail in the forthcoming sections. It must be noted that these are taken at summary levels.

- RVD Loyalty
- RVD Trend
- RVD Satisfaction

The suggested scoring for this is the overall score post a blood donation (either with a series of questions and the average value of the scores per individual) with the following scaling (1- No Satisfaction, 2- Low Satisfaction, 3- Average Satisfaction, 4- Good Satisfaction, 5- Excellent Satisfaction). Please note the data used is to be considered only for illustrative purposes.

7.4 Analysis

While extensive work has been done in the areas of identifying blood donor behaviour, we have enhanced the findings of the prior research and to the creation of new metrics that suggest and enable
decision making. The following are summary level metrics on RVD behaviour. In specific three new attributes have been introduced that can help assist in the effective campaign and overall donor management. These will help in the retaining of the first time donors [WB11]. The analysis of the loyalty, tendency and satisfaction index will help measure and identify trend patterns and facilitate conversion of first time donors to RVD.

- **Loyalty** means an RVD remains consistent to regular blood donation. This is tracked with the help of the Blood Donor Management System (BDMS) which help us record the persons donation. This is also derived over a range (one or more) milestone donorship drives. This is a summary level indicator. This is computed as the average of \( \text{RVDDuringMilestones-Timeframes} \) over the Average \( \text{DonorsDuringMilestonesTimeframes} \).

- **Tendency** is a measure of tendency pattern based on the donor behaviour. This is a pure statistical measure computed based on correlation between the \( \text{RVD During Milestones Timeframes} \) vs. \( \text{Donors During Milestones Timeframes} \).

- **Satisfaction index** is linked to the Blood donor management CRM that indicates the average rating feedback taken post the blood donation. This is computed as Average \( \text{SatisfactionIndex} \) across the milestones time-frames again at the summary level. This helps interpret the fuzzy areas around improving blood donor behaviour.
The RVD data is converted into a number of transformations to compute the three attributes. The following are some basis for the algorithm computation. The data is captured across multiple milestones at each location across a time span. These have to be facilitated by the BDMS systems and processes. The algorithm presented will use these data points and formula to derive the measure and trends.

**Loyalty Computation**

This measure is summarized at the location level. The computation is based on the following formula.

\[
\text{RVD Loyalty (Location L)} = \frac{\text{Average (RVD During Milestones Timeframes)}}{\text{Average (Donors During Milestones Timeframes)}}
\]  
(7.1)

The reporting systems captures this data from the Blood Donor Management System (BDMS) need to be configured for setting the required time-frames and milestones. These are to be configured as per the requirement of the implementation. The drill-down criteria is also configured as per requirement.

**Tendency Computation**

This measure is summarized at the location level. The computation is based on the following formula.

\[
\text{RVD Tendency (Location L)} = \text{Correlation (RVD During Milestones Timeframes, Donors During Milestones Timeframes)}
\]  
(7.2)
The reporting systems captures this data from the Blood Donor Management System (BDMS) need to be configured for setting the required time-frames and milestones. These are to be configured as per the requirement of the implementation. The drill-down criteria is also configured as per requirement.

**Satisfaction Computation**

This measure is summarized at the location level. The computation is based on the following formula.

\[
\text{RVD Satisfaction (Location L)} = \text{Average(Satisfaction Index Post The Milestones Timeframes)} (7.3)
\]

It must be noted that this measure is derived from the Customer Relationship Management (CRM) system. The satisfaction surveys are typically conducted post the donation (electronic or manual) by the donation center with feedback form the participants of the donation drive. These are captured into the CRM system and scored. The following is the score values for this measure. with the following scaling (1- No Satisfaction, 2- Low Satisfaction, 3- Average Satisfaction, 4- Good Satisfaction, 5- Excellent Satisfaction). Please note the data used is to be considered only for illustrative purposes.

The reporting systems captures this data from the Blood Donor Management System (BDMS) need to be configured for setting the required time-frames and milestones. These are to be configured as per the requirement of the implementation. The drill-down criteria is also configured as per requirement. In the case of this measurement an possible interface with CRM systems may be required (as required
by the implementation).

The suggested new scores are the computation of the RVD
tendency, RVD Loyalty and RVD Satisfaction index (linked to donor
behaviour and provided to the blood bank CRM which again get
rolled up at the level of drill down. It must be noted that this
study randomized the observation values used for computation of
these advanced attributes. The following flowchart indicates the
mechanism of computing the indicators.

Loyalty-Tendency-Satisfaction Scoring Algorithm

• Pre-requisite: Setup BDMS (time-frames and milestones), drill-
down configurations and CRM interfaces

• Step 1: Loop through each unique location L (latitude, lon-
gitude) and drill down (geographic, parameters) based on
requirement (such as state, district and city)

• Step 2: For each location L compute the average frequency,average
recency and total RVD count across the milestones and period
defined.

• Step 3: Calculation of Location level summary scores for the
advanced measurements Computation (Location L)

  • * The ComputeLoyalty (Location L) (formula 7.1)

  • * The ComputeTendency (Location L) (formula 7.2)

  • * The ComputeSatisfaction (Location L) (formula 7.3)
• Step 4: Plot this score in the chart with scores on the X-axis and locations on the Y-axis

Figure 7.1: Loyalty-Tendency-Satisfaction Scoring Flowchart
7.5 Results

The results of implementation of the advanced indicators to the test data provides the scores (figure 7.2). The scores will help identify the locations that need more focus. The drill down when customized to the specific implementation and needs will yield meaningful patterns for decision making as well as understanding the status.

The derived indicators location-wise help understand the key trends. The suggestive chart derived from the test data indicates the following patterns.

- Location L1 - High loyalty score, Lower tendency and satisfaction scores for RVD
- Location L2 - High loyalty score, Lower tendency and satisfaction scores for RVD
- Location L3 - High satisfaction score, Lower tendency and loyalty scores for RVD
• Location L4 - High satisfaction and tendency score, Lower loyalty scores for RVD

• Location L5 - High satisfaction and tendency score, Slightly lower loyalty scores for RVD

• Location L6 - High tendency score followed by satisfaction and loyalty scores for RVD

• Location L7 - High loyalty, tendency and satisfaction scores for RVD

These insights provide a data driven approach for healthcare policy makers to plan appropriate approaches for optimal management. These suggestive techniques can be customized for specific implementations (as per specific requirements of the blood center). When linked to real-time and realistic data the value addition is much more as these provide indicative time context patterns.
Table 7.1: Location-wise Recency Frequency and RVD details

<table>
<thead>
<tr>
<th>Location</th>
<th>Recency</th>
<th>Frequency</th>
<th>RVD</th>
<th>Loyalty</th>
<th>Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>9.71</td>
<td>5.75</td>
<td>6</td>
<td>0.35</td>
<td>-0.55</td>
</tr>
<tr>
<td>L2</td>
<td>9.47</td>
<td>5.82</td>
<td>3</td>
<td>0.28</td>
<td>-0.20</td>
</tr>
<tr>
<td>L3</td>
<td>13</td>
<td>6.22</td>
<td>1</td>
<td>0.31</td>
<td>0.07</td>
</tr>
<tr>
<td>L4</td>
<td>8.98</td>
<td>4.89</td>
<td>1</td>
<td>0.23</td>
<td>-0.28</td>
</tr>
<tr>
<td>L5</td>
<td>9.37</td>
<td>4.93</td>
<td>0</td>
<td>0.34</td>
<td>-0.77</td>
</tr>
<tr>
<td>L6</td>
<td>6.63</td>
<td>8.5</td>
<td>0</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>L7</td>
<td>9.55</td>
<td>5.48</td>
<td>2</td>
<td>0.24</td>
<td>-0.07</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.6 Conclusion

The ability to identify the critical patterns of blood donor behaviour has been demonstrated in this research study. The following provide some indicative results (table 7.2).

- Positive correlation between recency and RVD Tendency
- Negative correlation between recency and RVD Loyalty
- Positive correlation between frequency and RVD Loyalty

It must be noted these indicate a representation at certain time snapshot as-well as specific to the location and specifics of the milestones (specific to blood donation campaigns of the organization). The ability to review these key metrics provides decision makers and planners to effectively manage this function.

The research provides the ability to rank and score locations (figure 7.4). The score logic helps to understand the location specific performance (table 7.2). While it has been demonstrated in this manner in this research. It can be customized as per the requirement (figure 7.4).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Order</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Recency</td>
<td>Descending</td>
<td>Lower Value. Higher Score</td>
</tr>
<tr>
<td>Avg. Freq.</td>
<td>Ascending</td>
<td>Higher Value. Higher Score</td>
</tr>
<tr>
<td>Loyalty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tendency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This demonstrates a viable mechanism to manage blood donorship. In specific this helps address the optimized management of budget resources for blood donation management by providing critical insights for policy makers. The creation of dashboards (figure 7.5) with reports (basis these algorithms) will assist in proactive and data driven decision support of blood donor management.

<table>
<thead>
<tr>
<th>Location</th>
<th>AvgRecency</th>
<th>AvgFrequency</th>
<th>RVD</th>
<th>RVDLoyalty</th>
<th>RVDTendency</th>
<th>RVDsatisfactionIndex</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>L2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>L3</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>L4</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>L5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>L6</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>L7</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**Figure 7.4: Measurements and Score Logic**
Figure 7.5: Loyalty-Tendency-Satisfaction Dashboard
Future work will be focus on implementation of these algorithms and findings. Further enhancing these models to allow integration with blood donor management systems (BDMS) which include innovative ways of visualization (figure 7.6). The research provides a proof of concept basis approach to building an effective DM driven BDMS. The key advantage of this approach is that it allows the healthcare professionals to proactively work with various operational and process aspects of the BTS to provide and required levels of support. The suggested Data Mining Driven Blood Donor Management Systems (DM-BDMS) framework consists of the following key principles (figure 7.7).

- The BDMS and the CRM provide the transactional data to the
Blood Donor Analytics Database (BDAD)

- The BDAD contains the data required for the Blood Donor DM Models (BD-DMM)
- The BD-DMM implement the RVD, Loyalty, Tendency and Satisfaction computation models
- The BD-DMM models provide feedback to the BDAD
- The BDAD provides the feedback back to the base systems namely the BDMS and the CRM
- The BDAD in addition helps provide a knowledge base of patterns that can be reviewed and curated by subject matter experts
- The BDAD dashboards facilitate real-time and proactive push information
Figure 7.7: DM Driven BDMS Framework