www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 6 June 2017, Page No. 21679-21683
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i6.20

Using predictive modelling for forecasting blood donor response

Tejaswa Chaudhary, Gulzar Ahmad, Vikramjeet Mallick, Thirunavukkarasu K.
Student, School of Computer Science and Engineering, Galgotias University, Greater Noida, India
tejasw22@gmail.com
Student, School of Computer Science and Engineering, Galgotias University, Greater Noida, India
ahmad.gulzar8@gmail.com
Student, School of Computer Science and Engineering, Galgotias University, Greater Noida, India
vikramjeetmallick@gmail.com
Professor, School of Computer Science and Engineering, Galgotias University, Greater Noida, India
thiru.k@galgotiasuniversity.edu.in

Abstract: Blood and blood products are essential for medical treatment of all age groups. The primary source for blood products in the world is volunteer donors. Thus, donor recruitment and donor retention are vital factors for a blood bank to maintain its supply. We propose that developing a better understanding of donors' motivations to donate and personalizing their donation objectives would improve a blood bank's ability to secure a more robust supply of blood**.** The research paper puts forward a solution to predict whether a specific donor will donate blood in the coming month or not. This will help the blood banks to forecast their stock of various types of bloods and prepare accordingly.

Keywords: Predictive Modelling, Data Mining, Regression, RFM Model

1. Introduction

Health government agencies all over the world are continuously increasing their investment in information technology. The use of technology in health sector has given rise to the formation of huge EHR or electronic health records by various federal or private health organizations and also by individual doctors or physicians. This enormous increment in collection of data from medical bodies has enabled the health and welfare sector to harness the advantages of Big Data and Predictive Analytics to improve the quality and performance of their service. They can make methodical use of information acquired along with quantitative as well as quantitative analysis to make better decision and present improved results.

Analytics is being already used in several aspects of health care for support in clinical diagnosis decision making, remote monitoring of health along with medical resource allocation. The advent of big data in medical sciences has paved way for an ever-increasing requirement of informatics professionals who can bridge the gap between information and medical sciences.

2. Algorithms and Modules Used

2.1 Regression

In data analytics and regression can be used to define the association and connection between a scalar dependent value and a single or multiple explanatory/independent variables. When only a single explanatory/independent variable is used, the prediction can be made through a linear regression model. For more than one explanatory variable multiple linear regression is used.

In the regression models, the association are explained using predictor functions whose unknown model parameters are estimated from the present data.

In the solution proposed in this paper we will use frequency of donating blood as the scalar dependent value and use the latest month in which the blood was given as the independent variable. We will run these two variables through a linear regression model to predict whether the donor will donate blood in the future or not.

2.2 RFM Model

Various consumer driven companies use the RFM model (Recency, Frequency and Monetary) to predict quantitatively whether a certain customer will churn or not i.e. stay loyal to them and keep on buying their products or services or not. They examine how recently has the customer made a purchase, how repeatedly they made their purchases, and what is the magnitude of the amount they spent. This helps the company to single out customers who are about to switch their brand and entice them with lucrative offers to make them stay loyal. It is based on the marketing proverb that eighty percent of the business comes from twenty percent of the customers.

In our solution to predict whether a certain donor will donate blood in the upcoming month or not we apply the RFM model and substitute its original parameters with those relevant to donor statistics. The recency of product purchased is substituted with how recently has the donor donated blood, the frequency of purchases has been substituted with the frequency at which the donor donated blood and the money spent by the customer is replaced with the amount of blood donated (Here, the amount of blood of donated will depend on the frequency

of blood donation, as the amount of blood donated is fixed per donation).

3. Dataset

The dataset for the application purposes has been taken from the open database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan.
The dataset consists of 748 donors out of which are kept aside for 248 validations purposed and the remaining are used for the model implementation.
The dataset comprises of individual record of donors which include their last donation month, how many times they have donated and how much amount they have donated until now.

The dataset has been created from the actual database of a mobile blood bank bus over a period of three years. The bus went from universities to universities and offices to offices for obtaining donors. As the places, the bus visited were recurring after a certain amount of time (the time difference between the subsequent visit was equal to the time in which re-donation of blood was permissible).

4. Utility of the model

Often times it happens that a blood bank is unable to meet the requirement. This is due to them not able to forecast how many donors of what blood type will donate blood next month and will it be similar to this current month or not.
The model proposed predicts whether a specific donor will donate blood in the upcoming month or not. With this the blood banks can be assured that at least this amount will be available to them

5. Model Implementation

4.1 Regression

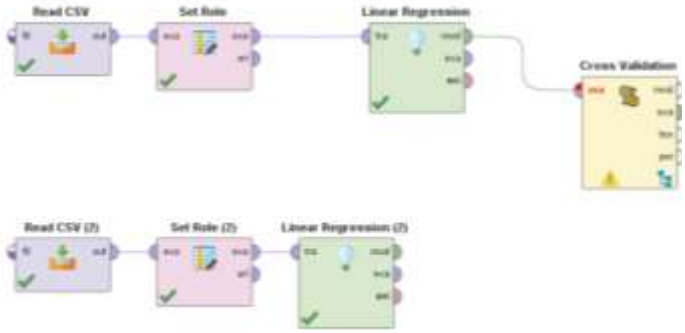

Figure 1Regression Model in Rapidminer

Total Data Elements – 748
Training Set – 500
Testing Set- 248

In our linear regression model, we use cross validation to check whether our model is giving correct results or not.

As the output for the regression model, a survey graph is given. The testing graph is almost the same to the training graph due to which we can conclude that the model is validated.

Interpretation of the given regression graph can be used as an indicator whether the donor will donate blood in the coming month or not.

4.2 RFM Model

Figure 2: RFM Model in Rapidminer

The RFM model analyses all the given information and passes it through the RFM function to get the RFM score. The score is the predictor whether the donor will give blood or not.

The RFM score is a binary output. If the output for a particular donor is 0 then it means that he/she will not donate blood but if it is 1 then it means that they will donate blood.

www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 6 June 2017, Page No. 21679-21683
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i6.20

6. Results

Figure 3: Stats of the initial Data

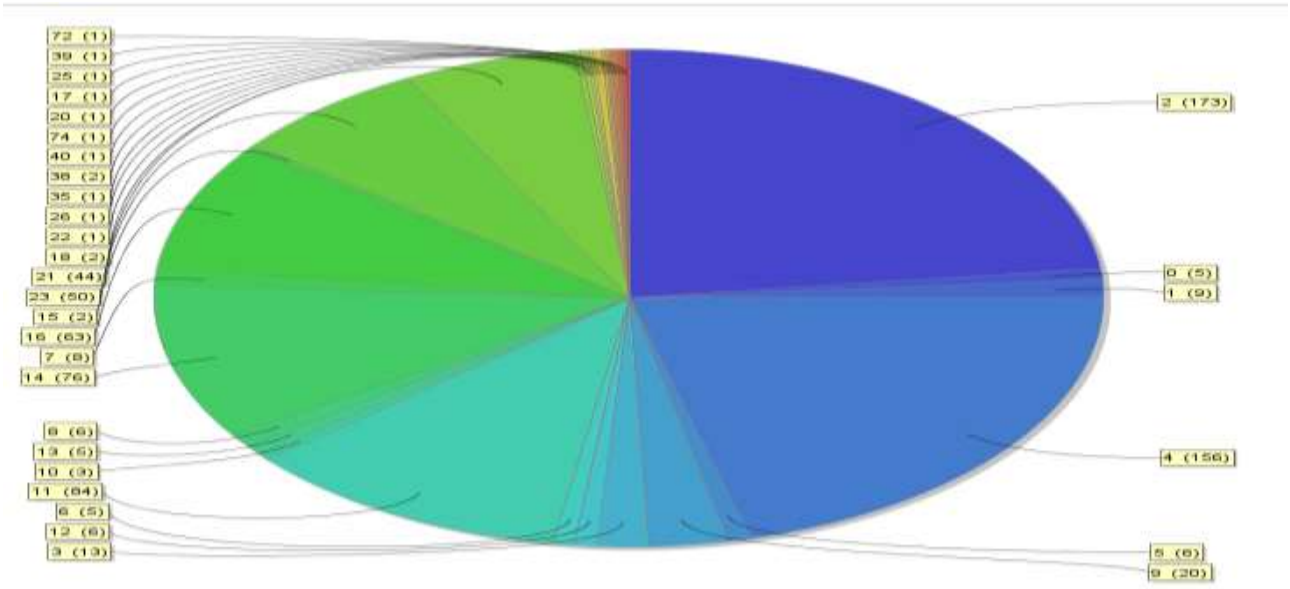

Figure 4: Frequency of Blood donation by various donors

DOI: 10.18535/ijecs/v6i6.20

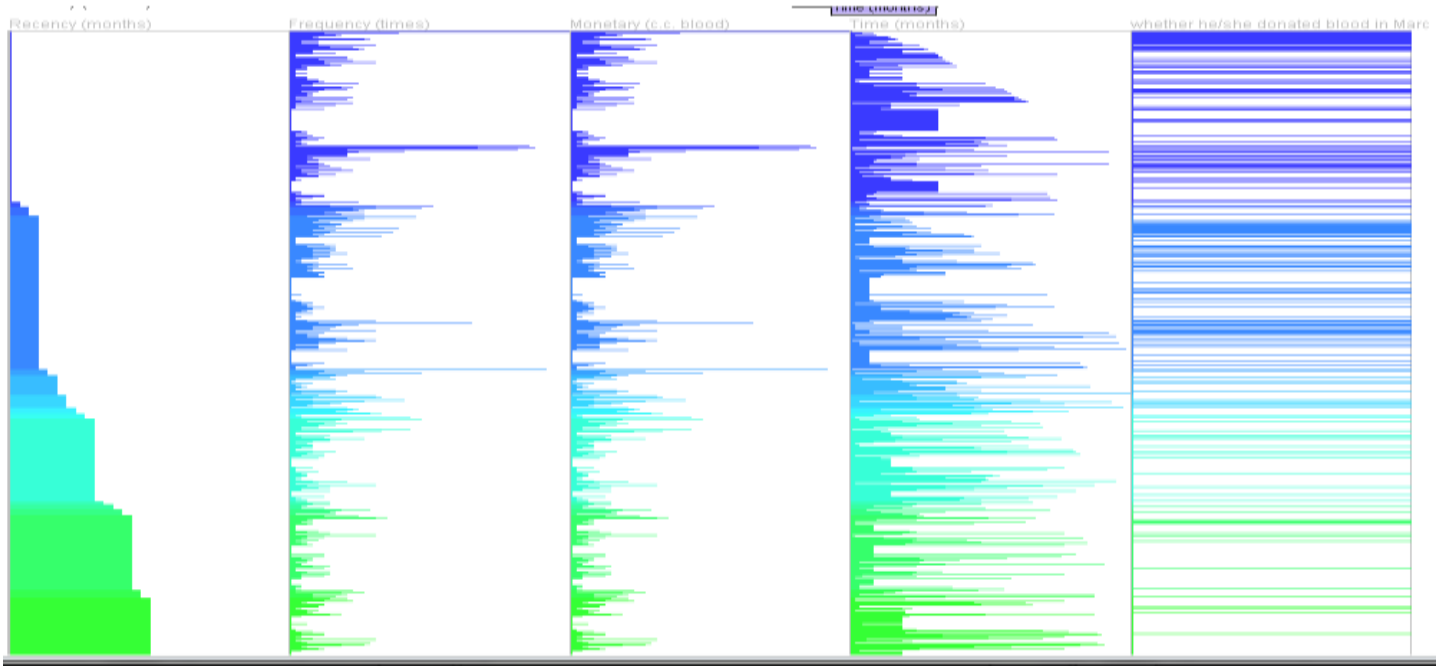

Figure 5: Survey Graph of various parameters

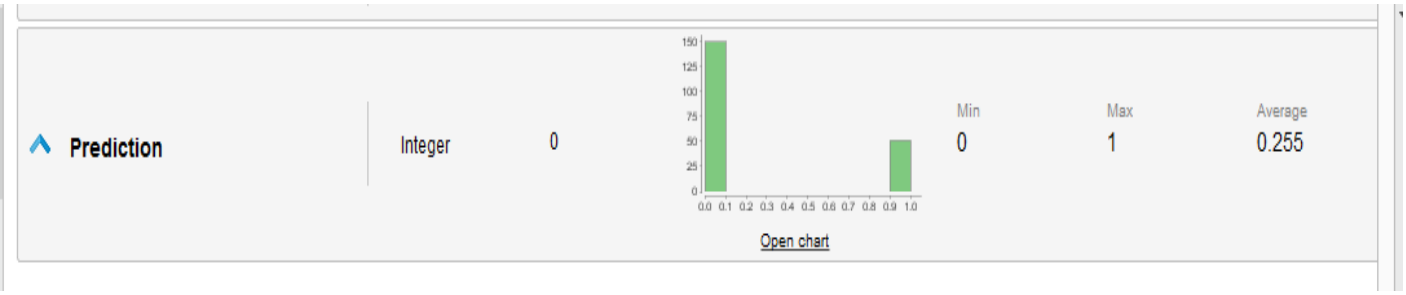

Figure 6: Prediction Results

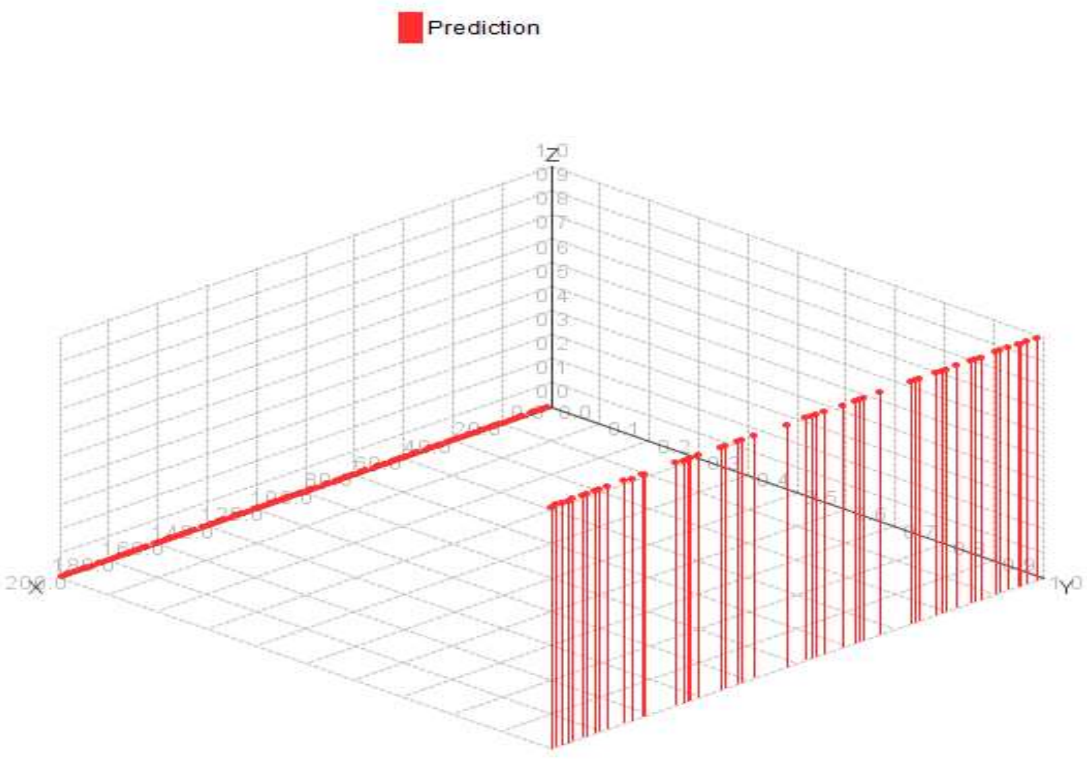

Figure7: Result Graph

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 6 June 2017, Page No. 21679-21683
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i6.20

7. Conclusion

Big data has drastically incremented the amount of information available to blood banks, they can use their vast database to be better prepare for future emergencies and save lives in the process.
This project aims to make proper and utmost use of the blood donated by donors. As a lot of patients die every year when they go through shortage of blood and are unable to procure it in right amount of time.

8. References

- DW Bates, S. S.-M. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs.
- J Soni, U. A. (n.d.). Predictive data mining for medical diagnosis: An overview. International Journal of Medicine.
- JW Bos, K. L. (2014). Private predictive analysis on encrypted medical data. Journal of biomedical informatics.
- M Geatz, R. R. (2003). System and method for sensing and evaluating physiological parameters and modeling an adaptable predictive analysis for symptoms management. US Patent App.
- S Soni, O. V. (2010). Using associative classifiers for predictive analysis in health care data mining. International Journal of Computer Applications.