A Predictive Analytics Framework for Blood Donor Classification

Kavita Pabreja, Maharaja Surajmal Institute, GGSIP University, Delhi, India

 https://orcid.org/0000-0001-9856-0900

Akanksha Bhasin, GGSIP University, Delhi, India

ABSTRACT

India faces numerous challenges to the meet ever-increasing demand of human blood so as to improve the health indicators across its rural and urban population. The gap between demand and supply can be fulfilled by increasing voluntary blood donations. Hence, it becomes important to understand the attitude of population towards blood donations. In this paper an effort has been made to identify features in order of their importance that affect the decision of a person to become a blood donor. This research uses extensive visualization techniques to get an insight into potential blood donor characteristics and then applies classification technique to classify youth of an Indian state university as donor or non-donor. The k-nearest neighbour classification algorithm discovers the relationship between attributes of blood donors and hence predicts the outcome. The important factors that dissuade potential donors from donating blood have been extracted that can be worked upon to meet the demand of blood to save human lives.

KEYWORDS

Blood Donor, Classification, Data Visualization, K-Nearest Neighbour, Lazy Learner Algorithm, Logistic Regression, Machine Learning, Random Feature Elimination

INTRODUCTION

Human blood is the precious constituent of life and there is no substitute for it. There has always been an acute shortage of human blood as far as a developing nation like India is concerned as stated by Verma et al. (2016). It is mentioned by Abolghasemi et al. (2009) that the rate of blood donations in developing countries is eighteen times lesser as compared to that of developed countries. Voluntary blood donations meet a significant portion of blood requirement in countries with higher income as explained by Nigatu and Demissie (2015). This non-remunerated donation has been considered as best and safest by Gharebhaghian, 2005 and Rahman et al. (2011).

A report on National Estimation of Blood Requirement in India has mentioned that the country faces many challenges in maintaining a sufficient supply of blood and its products. With an ever increase in Indian population augmented by advancement in clinical medicine, the demand of blood far outweighs its supply. This is also emphasized by Agrawal et al. (2013) and Benedict et al. (2012)

in their research reports. According to World Health Organization (WHO), every country should be able to provide safe and adequate blood to its needy population. It is also underlined by WHO that a country can meet its blood requirements if just 1 percent of its eligible population donates. According to report by Office of the Registrar General & Census Commissioner, India Census, approximately 50 percent of India's population is in the age group 18-65years which is the eligible age group for blood donation yet India fell short of 1.9million units of blood in the year 2017. Hence, it becomes essential for India as a nation to understand the factors that dissuade people from donating blood. With proper preparation, potential donors can be identified and registered with blood donation banks.

The precise aim of this piece of study is to search for realistic and convincing features in the youths' data that could be valuable for envisaging the probability of his/her becoming a blood donor. An effort has been made to categorize candidates into donors or non-donors class on the basis of their characteristics related to blood donation. This is the first time that real datasets related to students' views, sentiments and myths towards blood donations has been collected from students of a state university of India.

The organization of the paper is mentioned here: Literature review is described in Background section. Research methodology section explains about Data Collection, Data Pre-processing, Data Analysis, Feature Extraction, Machine Learning Algorithms used for Study, Model Evaluation and Prediction, Feature Ranking. At the end, conclusions are mentioned.

BACKGROUND

Various data mining techniques have been used extensively by researchers for classification, prediction, clustering, finding association and summarization tasks in the healthcare field. One of the unsupervised data mining techniques named k-means clustering, has been used to categorize the blood donors based on the gender, age, weight and blood group. The authors, Ramachandran *et. al.* (2011), have used the datasets from Indian Red Cross Society Blood Bank. A system has been developed by ChanLee and Cheng (2011) that uses classification and clustering algorithms to determine the variations in blood donation behaviour amongst the present donors and envisage their intents towards donation so as to understand various matters and to increase the voluntary blood donation frequency. The authors have applied clustering technique to create four groups and have found that the best accuracy is 0.783.

In order to understand the awareness and attitude of students of Semnan university of medical sciences, a descriptive analytical approach has been used by Majdabadi et al. (2018). It was found that a large number of students are not aware of blood donation and possess a negative attitude towards blood donation.

In order to help the humanity and save precious lives, a web-based system for maintaining records of blood donors has been created by Khan *et al.* (2009). The system registers the donors and keeps their record that has details of blood donors' blood groups, address for communication, and status of blood donation. This web-enabled system acts as an interface between donors and receptors. Similar web enabled systems have been developed and deployed by Arif et al. (2012) and Guangpeng et al. (2009). With the wide spread usage of mobile communication technologies, a few notification based systems have also been deployed by Singh et al. (2007), Rahman et al. (2011), Samsudinnet al. (2011) and Islam et al. (2013).

In a study by Hamouda *et al.* (2012), automatic red blood cell has been recognized and counted using image processing. Decision tree has been used to classify Red Blood Cells that has classified the data with an accuracy of 97%. Real datasets collected from an Electronic Data Processing wing of a blood bank has been classified using J48 algorithm by Sharma and Gupta (2012) that can facilitate the blood bank in-charge to make suitable decisions quicker and more accurate. There is a study by Mostafa (2009) where Intelligent data modelling techniques have been used in Egypt to examine the impact of demographic, perceptive and psychological factors on blood donations. The author has observed that there are five factors that are important for understanding blood donors' behaviour, *viz.*

Altruistic values, knowledge of blood donation, intent to donate blood, perceived risks of donation of blood, and attitude towards blood donation. A framework for the predictors for behaviour of established Australian blood donors has been determined by Masser *et al.* (2009).

Rajput *et al.* (2009) have stated that it is a great challenge to utilize data mining algorithms in the fields of healthcare and medicine. A report by Government of India (2007) has suggested that voluntary non-remunerated regular blood donations are the safest. The strategy of Indian government focuses on motivating non-remunerated blood donors and it emphasizes to maintain good epidemiological data on the occurrence of infectious markers in the general population. One of the main hindrances for blood donations are risks associated with the process as explained by Tscheulin and Lindenmeier (2005) that mainly includes fear of infection.

Qualitative studies have been used by Ferguson and Chandler (2005) to express that blood-donors depict their behaviour using Trans Theoretical Model. Schlumpf et al. (2007) have done extensive study based on a questionnaire filled in by approximately 8000 active donors. The possibility of return of a current donor within next 12 months has been explored using logistic regression.

The prediction of blood donor using age and blood group has been done by Sharma and Gupta (2012). The authors have made use of WEKA tool for data mining. A data mining system based on clustering and classification has been developed by Chan-Lee and Cheng (2011) [7.I] in order to understand the behaviour of blood donors.

From all these studies, it is clear that it is very important to remove myths by educating people and also to identify donors so that the blood banks and other voluntary organizations chalk out a strategy for organizing blood donation camps. By applying classification technique, the potential donors can be identified and the important factors that dissuade eligible donors from donating blood can be extracted. This research work is based on data of an Indian university's students so as to understand their knowledge, attitude and psychology towards blood donations like their fears, myths, risks while becoming a blood donor. This paper uses machine learning algorithms viz. k-nearest neighbour and logistic regression to classify potential donors as donor or non-donor. It also makes use of feature extraction to find and rank important features that play a significant role for a person to become blood donor. The ultimate objective is to motivate such eligible people to donate blood regularly so that many human lives can be saved.

RESEARCH METHODOLOGY

The population used for the research belongs to Generation Z (born between 1995 and 2015). These are the students of Undergraduate programmes of a Delhi state university, India. The system framework showing all steps of research in order to perform predictive analytics of blood donors, is shown in Figure 1.
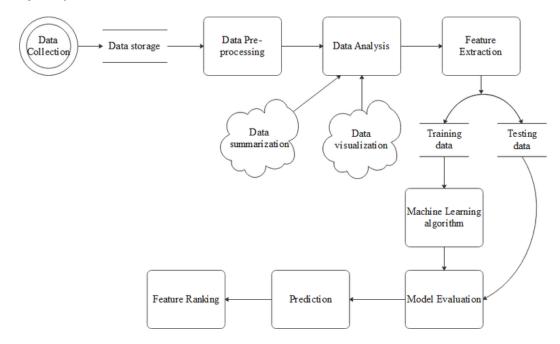
Online Collection of Data

Data has been collected by using questionnaire developed in google forms by students of an undergraduate programme of a Delhi state university in order to complete their major project. Approximately 500 students of ten different colleges, pursuing undergraduate programmes have been surveyed and responses have been gathered. Convenience sampling technique has been used and hence the selection of the participants was non-random and voluntary.

The questionnaire is based on personal attributes, intention towards blood donation, myths related to blood donation, risks associated with blood donation and perceived belief of probable donors before taking decision on blood donation. There is a total of 20 questions that includes one question describing the class of blood donor (i.e. whether the person is willing to become a blood donor or not). The response to nineteen questions is on Likert scale in which the respondents were asked to select the choice that suited them the most. The choices are Definitely yes, Probably yes, Maybe

Figure 1. System framework for research

yes, Probably no and Definitely no. The complete description of the questionnaire is mentioned in Table 10, in the Appendix.

Data Pre-Processing

Out of 20 questions, nineteen have responses on Likert scale. One-hot encoding which is a dummification technique, has been used on these nineteen features. This encoding is basically the representation of categorical variables as binary vectors. These categorical values are first mapped to integer values. Each integer value is then represented as a binary vector that is all 0s (except the index of the integer which is marked as 1). This transformation is required so as to prepare datasets for feeding to an appropriate classification algorithm in Python. The output attribute is "Willingness to become donor" that can take value *viz.* Yes / No. "Yes" has been transformed to "1" and "No" has been converted to "0".

Data Analysis

Data Summarization

A total of 448 participants responded and their frequency distribution on basis of willingness to donate blood, blood groups and religion are shown in Table 1, Table 2 and Table 3 respectively.

Table 1. Distribution of willingness to donate blood as reported by participants

Willingness to donate blood	Count	Percentage
Yes	258	57.58%
No	190	42.42%

Table 2. Distribution of blood group as reported by participants

Blood group	Count	Percentage
A+	82	18.30%
A-	14	3.13%
AB+	57	12.72%
AB-	9	2.01%
B+	186	41.52%
B-	15	3.35%
O+	76	16.96%
O-	9	2.01%

Table 3. Distribution of religion as reported by participants

Religion	Count	Percentage
Buddhism	11	2.46%
Christian	10	2.23%
Hindu	346	77.23%
Jainism	24	5.36%
Muslim	19	4.24%
Sikh	34	7.59%
Other	4	0.89%

Data Visualization

Various visualization and numerical calculations libraries of Python have been used to understand the attitude of respondents towards blood donations. These libraries are seaborn and matplotlib.pyplot for generating barplots and numpy for numerical calculations like grouping the participants on the basis of their response on likert scale. Bar plot showing percentage of respondents in each of the five options to a particular question, has been generated. There was a total of 17 such questions to understand the characteristics of participants and hence 17 such graphs are generated as shown in Table 4.
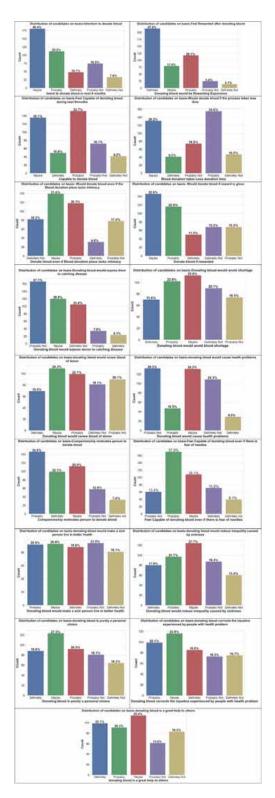
Feature Extraction

The data visualization bar plots have been interpreted so as to select only those features for input to machine learning algorithm for classification. A few of the bar plots do not show much of variation and the distribution of respondents in each of the five categories on likert scale is of similar nature, so these features were removed before applying machine learning algorithms. The removed features are donating blood is purely a personal choice, donating blood would renew blood of donor, donating blood would avoid blood shortage.

Machine Learning Algorithms Used for Study

Two popular machine learning algorithms *viz.* a lazy learner classifier (K-Nearest Neighbor) and logistic regression have been used on Spyder which is a powerful scientific environment written in Python.

Table 4. Distribution of participants in each of the five options of Likert scale, as per their response to each of the 19 questions

K-Nearest Neighbor (K-NN) is an algorithm for classification which memorizes the training data first and when presented with a testing record, it looks for similarity to the memorized training records. Whichever training record is most similar to the test case, that class is assigned to the test tuple. As explained by authors in (Peng et al., 2009) the benefit in using a lazy learning algorithm is that there is local approximation of target function for each query posted to the classifier. This leads to solving many queries in an organized and easy manner. K-NN classifier applies an incremental approach wherein the input comprises a set of attribute-value pairs, as described by Witten Eibe (2011). There is one attribute that corresponds to the class of tuple and other attributes are used as predictors.

Logistic Regression is a very popular machine learning technique based on statistics. It is a type of regression analysis method to apply when the dependent variable is dichotomous (binary). The logistic regression is a predictive analysis technique that uses a logistic function. As explained by Hosmer and Lemeshow (2000), it is used to describe the relationship between one dependent binary variable and one or more ratio-scaled, nominal, interval or ordinal independent variables.

For experimentation, K-NN and Logistic regression have been used for classifying the participants as blood donor or non-donor. The responses have been split into two parts, *viz.* training and testing. The training of the machine has been done with 70% of records and rest 30% have been used for testing.

Model Evaluation and Prediction

The machine learning algorithm for classification would predict the output class of a student as either Blood donor (Positive class) or Non-donor (Negative class). There are only four categories, given below, that any student X could end up with:

- True positive (TP): Prediction is Donor and X is actually a Donor.
- True negative (TN): Prediction is Non-Donor and X is actually a Non-Donor
- False positive (FP): Prediction is Donor but X is actually a Non-Donor, so it is a false alarm.
- False negative (FN): Prediction is Non-Donor but X is actually a donor, again a wrong prediction.

These four cases in confusion matrix are shown in Table 5.

Using the confusion matrix, a number of performance metrics have been calculated in Python. These metrics are explained below.

Accuracy

It is the ratio of the correctly labelled class to the entire collection of classes:

$$Accuracy = \left(TP + TN\right) / \left(TP + FP + FN + TN\right) \tag{1}$$

Precision

Precision is the ratio of the correctly predicted positive labelled records by the algorithm to all positive labelled records including wrongly labelled also:

Table 5. Confusion matrix

		Predicted class	
		Non-donor	Donor
Actual class	**Non-donor**	TN	FP
	Donor	FN	TP

$$Precision = TP / \left(TP + FP\right) \tag{2}$$

Recall (Sensitivity)

Recall is the ratio of the correctly predicted positive labelled records by the algorithm to all who are actually positive in reality:

$$Recall = TP / \left(TP + FN\right) \tag{3}$$

F1-Score (F-Measure)

F1 Score takes into account both precision and recall. It is the harmonic mean of the precision and recall. It is a good indicator of performance of classifier when there is uneven class distribution:

$$F1 Score = 2 * \left(Recall * Precision\right) / \left(Recall + Precision\right) \tag{4}$$

Specificity

Specificity is the ratio of the correctly predicted negative labelled records by the algorithm to all who are actually negative in reality:

$$Specificity = TN / \left(TN + FP\right) \tag{5}$$

K-NN Algorithm

The value of k has been varied from 1 to 10 in order to find the maximum value of correctly classified records. It is found that the best classification accuracy is when k=8. The corresponding confusion matrix is shown in Table 6. Using this confusion matrix, the calculation of mentioned performance metrics has been done and is shown in Table 8.

Logistic Regression

By applying this machine learning algorithm, the predicted class vs. actual class data has been shown in confusion matrix in Table 7. Using this confusion matrix, the calculation of mentioned performance metrics has been done and is shown in Table 8. As it is evident from Table 8, the K-NN algorithm with k=8 has outperformed the logistic regression in all the performance metrics.

Feature Ranking

The task of determining the important features (independent) that are greatly affecting the decision of a student to be a blood donor has also been done. For this purpose, Recursive Feature Elimination

Table 6. Confusion matrix for K-NN algorithm

		Predicted class	
		Non-donor	Donor
Actual class	**Non-donor**	TN = 33	FP = 24
	Donor	FN = 16	TP = 62

Table 7. Confusion matrix for Logistic Regression algorithm

		Predicted class	
		Non-donor	**Donor**
Actual class	**Non-donor**	TN = 31	FP = 26
	Donor	FN = 17	TP = 61

Table 8. Comparison of performance measures for K-NN and Logistic Regression Algorithms

Performance measure	K-NN algorithm	Logistic Regression
Accuracy	0.7037	0.6815
Precision	0.7209	0.7011
Recall (Sensitivity)	0.7949	0.7821
F1-score (F-Measure)	0.7561	0.7394
Specificity	0.5789	0.5439

(RFE) method that generates feature importance ranking, has been used. RFE is a feature selection approach that gives the ranking of the features according to their importance for determination of the dependant variable. It uses the model accuracy to find which attributes (and combination of attributes) contribute the most to predicting the dependent attribute. There are many benefits of using RFE viz. reduction of overfitting, reduction of training time and improvement in accuracy of the model. The scikit-learn Python library provides this method. RFE model is created by using logistic regression classifier as a base model. This model has listed important features as per their ranks as shown in Table 9.

CONCLUSION

In this paper, we have experimented with real datasets of students of Undergraduate programmes of an Indian university. A total of seventeen questions related to students' views, awareness, sentiments and myths pertaining to blood donations were asked using online questionnaire. Data visualization technique has been used to remove those features that are not contributing much towards decision of a participant to donate blood. Following this, two machine learning algorithms *viz.* K-Nearest Neighbor and Logistic Regression, for classification have been used to label students as donors

Table 9. Names of features as per their ranks that are main deciding factors for a person to donate blood

Rank	Feature Name
First	Would donate blood if the process of donation takes less time.
	Would donate blood as donating blood would be a rewarding experience.
	Would donate blood if accompanied by friend, family member or colleague.
	Would not like to donate blood if the place lacks intimacy.
Second	Would donate blood if it reduces inequality caused by sickness.
Third	Would donate blood if blood donation by me permits a sick person live with a better health.

or non-donors. K-NN algorithm was executed by varying value of k from 1 to 10 and the best classification accuracy was obtained with value of k equal to 8. After performing comparison of various performance metrics, it was found that the K-NN classifier has demonstrated convincing results with an Accuracy of 0.7027, Precision equal to 0.7209, Sensitivity value 0.7949, F1-score equal to 0.7561 and Specificity value 0.5789.

This study has provided the ability to identify important factors that influence the decision of youth to donate blood. These factors are time taken by blood donation process, companionship of a friend or family member while donating blood, availability of intimate place for blood donation and feeling of being rewarded. With the knowledge of these important determinants, the blood donation services can come up with newer and more efficient strategies that would increase the number of donors. This identification of probable donors would help blood banks and voluntary organizations plan in advance for the organization of blood donation camps. Also, the participants predicted as Non-donors can be motivated and the factors that restrict them from donation can be worked upon. Hence the significant gap between demand and availability of blood in India can be reduced by better management and collection of blood.

REFERENCES

Abolghasemi, H., Maghsudlu, M., Kafi-Abad, S. A., & Cheraghali, A. (2009). Introduction to Iranian blood transfusion organization and blood safety in Iran. *Iranian Journal of Public Health*, *38*(1), 82–87.

Agrawal, A., Tiwari, A. K., Ahuja, A., & Kalra, R. (2013). Knowledge, attitude and practices of people towards voluntary blood donation in Uttarakhand. *Asian Journal of Transfusion Science*, *7*(1), 59–62. doi:10.4103/0973-6247.106740 PMID:23559768

Arif, M., Sreevas, S., Nafseer, K., & Rahul, R. (2012). Automated online Blood bank database. *Annual IEEE India Conference*.

Benedict, N., Usimenahon, A., Alexander, N. I., & Isi, A. (2012). Knowledge, attitude and practice of voluntary blood donation among physicians in a tertiary health facility of a developing country. *International Journal of Blood Transfusion and Immunohematology*, *2*, 4–10. doi:10.5348/ijbti-2012-7-OA-2

ChanLee, W., & Cheng, B.W. (2011). An Intelligent system for improving performance of blood donation. *Journal of Quality*, *18*(2), 173–178.

Ferguson, E., & Chandler, S. (2005). A stage model of blood donor behaviour: Assessing volunteer behaviour. *Journal of Health Psychology*, *10*(3), 359–372. doi:10.1177/1359105305051423 PMID:15857868

Government of India. (2007). *Voluntary blood donation programme*. http://www.nacoonline.org/upload/Final%20 Publications/Blood%20Safety/voluntary%20blood%20do nation.pdf

Guangpeng, L., Zhongwen, G., Song, X., & Wenli, P. (2009). Web-based real-time monitoring system on cold chain of blood. *IEEE Instrumentation and Measurement Technology Conference*. doi:10.1109/IMTC.2009.5168655

Hamouda, A., Khedr, A. Y., & Ramadan, R. A. (2012). Automated Red Blood Cells counting. *International Journal of Computational Science*, *1*(2), 13–16.

Hosmer, D. W. & Lemeshow, S. (2000). Applied Logistic Regression (2nd ed.). John Wiley & Sons, Inc.

Islam, A., Ahmed, N., Hasan, K., & Jubayer, M. (2013). mHealth: Blood donation service in Bangladesh. *International Conference on Informatics, Electronics and Vision (ICIEV)*. doi:10.1109/ICIEV.2013.6572594

Khan, A.R. (2009). Web based Information System for Blood Donation. *International Journal of Digital Content Technology and Its Applications, 3*(2), 137-142.

Majdabadi, H. A., Kahouei, M., Taslimi, S., & Langari, M. (2018). Awareness of and attitude towards blood donation in students at the Semnan University of Medical Sciences. *Journal of Electronic Physician*, *10*(5), 6821–6828. doi:10.19082/68 PMID:29997767

Masser, B. M., White, K. M., Hyde, M. K., Terry, D. J., & Robinson, N. G. (2009). Predicting blood donation intentions and behaviour among Australian blood donors: Testing an extended theory of planned behaviour model. *Transfusion*, *49*(2), 320–329. doi:10.1111/j.1537-2995.2008.01981.x PMID:19040598

Mostafa, M. M. (2009). Profiling blood donors in Egypt: A neural network analysis. *Expert Systems with Applications*, *36*(3), 5031–5038. doi:10.1016/j.eswa.2008.06.048

Nigatu, A., & Demissie, D. B. (2015). Knowledge, Attitude and Practice on Voluntary Blood Donation and Associated Factors among Ambo University Regular Students, Ambo Town, Ethiopia. *Journal of Community Medicine & Health Education*.

Peng, Y., Jianyong, Z., & Yumhong, X. (2009). Lazy learner text categorization algorithm based on embedded feature selection. *Journal of Systems Engineering and Electronics*, *20*(3), 651–659.

Rahman, M. S., Akter, K. A., Hossain, S., Basak, A., & 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*. doi:10.1109/WAINA.2011.115

Rajput, A., Aharwal, R. P., Chandel, N., Solanki, D.S., & Soni, R. (2009). Approaches of Classifications to Policy of Analysis of Medical Data. *International Journal of Computer Science and Network Security, 9*(11), 1-9.

Ramachandran, P., Girija, N., & Bhuvaneswari, T. (2011). Classifying blood donors using data mining techniques. *International Journal of Computer Science and Engineering Technology*, *1*(1), 10–13.

Samsudin, N., Khalid, S. A., Yusoff, A. M., Ihkasan, M., & Senin, Z. (2011). Procedure automation with immediate user notification: A case study. *IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA)*. doi:10.1109/ISBEIA.2011.6088816

Schlumpf, K. S., Glynn, S. A., Schreiber, G. B., Wright, D. J., & Randolph, S. W. (2007). Factors influencing donor return. *Transfusion*, *48*, 264–272. PMID:18005325

Sharma, A. & Gupta, P.C. (2012). Predicting the number of blood donors through their age and blood group by using data mining tool. *International Journal of Communication and Computer Technologies, 1*(2).

Singh, R., Bhargava, P., & Kain, S. (2007). Smart Phones to the Rescue: The Virtual Blood Bank Project. *IEEE Pervasive Computing*, *6*(4), 86–89. doi:10.1109/MPRV.2007.87

Tscheulin, D., & Lindenmeier, J. (2005). The willingness to donate blood: An empirical analysis of socio-demographic and motivation-related determinants. *Health Services Management Research*, *18*(3), 165–174. doi:10.1258/0951484054572547 PMID:16102245

Verma, S., Sharma, R. K., Sharma, M., & Pugazhendi, S. (2016). Voluntary Blood Donation: Attitude and Practice among Indian Adults. *Journal of Community Medicine & Health Education*, *6*(3). Advance online publication. doi:10.4172/2161-0711.1000436

Witten, I. H., & Eibe, F. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). The Morgan Kaufmann Series in Data Management Systems.

APPENDIX: DESCRIPTION OF QUESTIONNAIRE

Table 10. Description of the questionnaire

S.No	Variable	Description	Collected Data/ Response
1	BGroup	Blood Group	○ A+ ○ A- ○ AB+ ○ AB- ○ B+ ○ B- ○ O+ ○ O-
2	Religion	Religion	○ Buddhism ○ Christian ○ Hindu ○ Jainism ○ Muslim ○ Sikh ○ Other
3	Intend_to_give	Intend to give blood in next 6 months	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
4	Rewarding_exp	For me, donating blood during the 6 months would be a Rewarding Experience	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
5	Capability	I am eligible to donate blood during the next 6 months	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
6	Less_donation_time	If the process of donating blood took less time, I would donate blood	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
7	Intimacy_level	I would agree to donate blood even if blood donations takes place, in a place, which lacks intimacy	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
8	Given_a_reward	If I were given a reward, I would donate blood	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
9	Catch_a_disease	If I were to give blood, this would expose me to catching a disease	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
10	Avoid_shortage	If I were to donate blood, this would help avoid blood shortage	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
11	Renew_my_blood	If I were to donate blood, this would help me to renew my blood	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
12	Cause_health_problems	If I were to donate blood, this could cause me health problems	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
13	Companionship	If I were accompanied by a friend, a family member or a colleague, I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
14	Trypanophobia	I would be capable of donating blood, even if I were afraid of needles or fainting.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
15	Improved_sick_life	If blood donation by me permits a sick person live in a better health, I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
16	Reduced_inequality	If blood donation by me reduces inequality caused by sickness, I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
17	Personal_choice	If blood donation is purely my personal choice, I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
18	Injustice_by_health	If blood donation by me corrects the injustice experienced by people with a health problem that necessitates a blood transfusion, I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
19	Help_people	If blood donation by me helps my fellow men/women I would donate blood.	○ Definitely ○ Probably ○ Maybe ○ Probably not ○ Definitely not
20	Blood_Donor	Willing to become a blood donor	○ Yes ○ No

Kavita Pabreja is working as Associate Professor in the Department of Computer Applications at Maharaja Surajmal Institute, an affiliate of GGS Indraprastha University, New Delhi, India. She received her Ph.D. in Computer Science from Birla Institute of Technology & Science, Pilani. She holds more than 25 years of experience which includes teaching experience of 20 years and industry experience of over 5 years with Indian as well as USA MNC. She has authored four books, two with International publishers - "Application of Artificial Intelligence tools – Impact on Weather Prediction", "Object-Oriented Programming Using C++", and two with National publishers- "Learning Visual Basic.Net Programming", "Front End Design Tool VB.NET". She has contributed over sixty papers in International Journals / Book/ International conferences of repute. She is a recipient of a Best Research Paper Award from EMC data storage systems and Amity for All India Competition in Data Research; another Best Research Paper award for developing An Android Application to illustrate the Perspective of Community on Blood Donation at National Conference organized by IITM, affiliate of GGSIP University, India.

Akanksha Bhasin is currently a Post Graduate student, pursuing Masters of Computer Applications in Software Engineering at University School of Information, Communication and Technology, GGSIP University. She completed her graduation in the field of Computer Applications from Maharaja Surajmal Institute, an affiliate of GGS Indraprastha University, Delhi. She received Academic Excellence Award from MSI as well as Exemplary Performance Award from GGSIP University for the same for her outstanding academic achievements. Moreover, being a data science enthusiast, she published numerous research papers and has been conferred with Best Research Paper award for developing "BloodMate- An Android Application To Illustrate the Perspective of Community on Blood Donation" at National Conference organized by IITM, an affiliate of GGSIP University, Delhi.