

Predicting Return Donor and Analyzing Blood Donation Time Series using Data Mining Techniques

Anfal Saad Alkahtani¹, Musfira Jilani²

Collage of Information and Computer Sciences

Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia¹

School of Computing, Dublin City University, Dublin, Ireland²

Abstract—Since blood centers in most countries typically rely on volunteer donors to meet the hospitals' needs, donor retention is critical for blood banks. Identifying regular donors is critical for the advance planning of blood banks to guarantee a stable blood supply. In this research, donors' data was collected from a Saudi blood bank from 2017 to 2018. Machine learning algorithms such as logistic regression (LG), random forest (RF) and support vector classifier (SVC) were applied to develop and evaluate models for classifying blood donors as return and non-return donors. The natural imbalance of the donors' distribution required extra attention and considerations to produce classifiers with good performance. Thus, over-SMOTE sampling was tested. Experiments of different classifiers showed very similar performance results. In addition to the donors return classification, a time series analysis on the donors dataset was also considered to find any seasonal variations that could be captured and delivered to blood banks for better planning and decision making. After aggregating the donation count by month, results showed that the number of donations each year was stable except for two discovered drops in June and September, which for the two observed years coincided with two religious periods: Fasting and Performing Hajj.

Keywords—Classification; machine learning; time series analysis; blood donation

I. INTRODUCTION

Blood donation is an integral and essential part of the healthcare system. Without blood banks, many of the medical procedures that we otherwise take for granted could not take place. The modern lifestyle, ever-increasing mobility and accompanying higher accident rates, and incidences of natural and human-made disasters (such as wars, earthquakes, etc.) have led to an ever rising demand for blood transfusions [1]. A constant supply of blood is needed to help ensure that hospitals have access to enough blood to meet their current and future needs. One of the most important factors for a stable supply is the retention of donors, as return donors allow blood banks to not spend additional efforts and resources looking for new ones [2].

Kingdom of Saudi Arabia (KSA) is a large country with a total population of 33413660 inhabitants [3]. According to the Ministry of Transportation, 13221 car accidents were recorded in 2018 [4]. As a result, more blood banks in all regions were highly needed. Despite several campaigns aimed at promoting voluntary blood donations [5], statistics show that KSA is lacking the adequate number of volunteer blood donors [6].

In order to assist in overcoming these challenges, research on donors' data is critical. Machine learning and data mining methods can be applied to analyze the behavior of donors, allowing blood banks to build binary classification models for the return prediction of donors. However, in a real dataset of donors, there are few return donors compared with a high number of non-return donors. This imbalance in the donors' dataset introduces another problem to the binary classification, a condition where the model will consider too highly the majority class of non-return donor and ignores the minority class of return donor (even though the return donors are the primary class of interest). Thus, special considerations are required. Moreover, exploring seasonal variations or trends can help blood banks in planning and decision making. Time series analysis and forecasting methods are helpful to discover such knowledge.

This study was conducted on a real data set collected by the author from a public hospital in Saudi Arabia. Moreover, this is the first study applied in Saudi Arabia in the domain of data mining applications on blood donors.

In addition to the current section, this paper is divided into six sections. Section 2 starts with an overview of the history of blood donation and its system, then presents the previous works in return donors predicting and blood donation time series analysis. The data collection is shown in Section 3. Next, Section 4 describes all methodologies used in the study. Then, Section 5 presents the discussion. Finally, Section 6 summarizes the whole research in the conclusion.

II. LITERATURE REVIEW

A. Blood Donation System

The blood supply chain in general consists of three main roles: blood donors, transfusion agencies (such as hospitals), and blood centers (which attempt to coordinate the balance between supply and demand). The system can be described in three stages as shown in Fig. 1.

Stage 1: Blood Collection

Blood centers invite donors to donate blood in a number of different ways, such as recruitment campaigns, direct mailings and phone calls. The donor can be a volunteer, a replacement (a patient's relative or friend), or a paid donor. The blood collection could be for whole-blood, plasma, or platelet. The collection process is performed in government institutions, companies, and hospitals [7].

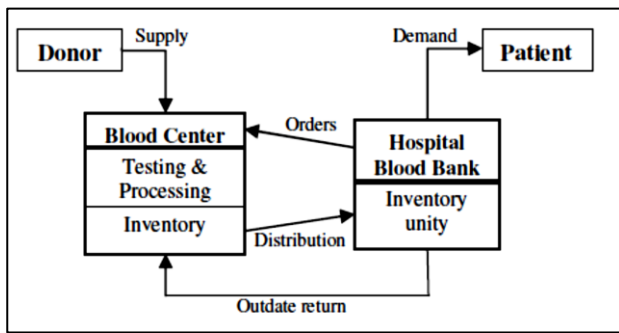


Fig. 1. Blood Supply Chain [7].

Stage 2: Blood test and processing

After collection, the blood centers apply serological and immunohematological tests on the collected blood bags. One of these tests is to determine the blood group of each bag, which can be one of the four different blood group (A, B, AB, and O) and is either Rh-positive or Rh-negative. Then, the bags are sent to processing units to extract and store blood components [7].

Stage 3: Blood distribution

Requests from hospitals to fill their blood bank needs are met by centers releasing and distributing blood components [7].

B. Blood Donation around the World and in Saudi Arabia

Blood centers all over the world are suffering a high shortage in the blood supply. Moreover, the demand for blood replacement is continually increasing in all countries [8]. The American Red Cross states that there is someone in need of blood every two seconds. However, the blood centers rely on volunteer donors, and most do not return to donate again. Unfortunately, not all donors are eligible, since a lot of donated blood bags are rejected after the test stage [9]. Another factor causing a decline of donation is that many returning donors are aging, whereas younger donors are rare [8].

Although the United States has a strong coordinated effort between the government and the Red Cross, the blood supply remains below the demand. In the U.S., the daily need of blood donors is around 44000 [10]. According to the Blood Transfusion Service in Northern Ireland, the yearly need of new donors is 10000 donors [10]. A study by Gibbs and Corcoran reported that “80% of developing countries depend totally or partially on replacement donors, 15% on voluntary/non-remunerated and 25% on paid donations” [9].

The KSA is a large country with a total population of 33413660 inhabitants [3]. According to the Ministry of Transportation, 13221 car accidents were recorded in 2018 [4]. As a result, many blood banks in all the regions are highly needed. At the Fourth World Medical Conference on Blood Transfusion organized at Prince Sultan bin Abdul-Aziz Center for Science and Technology (SITEC), the assistant undersecretary of the Ministry of Health for Laboratories and Blood Banks stated that there are a series of 6 central blood banks and 18 major blood banks provided for health services around the Kingdom. He stated that 60% of the need in the

blood banks in the Kingdom is usually covered by the donation of relatives and friends [6]. The voluntary donation of blood meets only 40% of the total need, which is too low, as he would like to see that closer to 100% [6].

The cycle of blood donation in the KSA is started by the donor and ends with the transfusion to the patient. It is essentially a hospital-based procedure, managed by the blood banks of the individual hospitals. Thirty years ago, importing a blood supply was stopped, and the Kingdom decided to depend entirely on local blood donations [11]. Currently, the source of blood donations is a combination of mostly replacement donors and a growing number of voluntary donors. The latter source is expanding rapidly through better operations by blood bank managers [11]. Previously, participation in blood donation was less than satisfactory among the Saudi public. This was probably due to a poor awareness of the importance of blood donation. However, this is changing as a result of better communications between the hospital blood banks and the general public [12] [13].

C. Toward Blood Bank Analysis using Data Mining Techniques

As a result of the increased global demand for blood, there is a serious need to keep an adequate supply of blood that is readily available [7]. Finding a way to recruit new donors and encourage previous donors to return is a major challenge for blood centers. In order to address this challenge, many studies have been conducted to apply different data mining tasks in the blood donation field. For instance, predicting return donors and forecasting the number of donors over a short time interval. This section reviews some of the most relevant research regarding these tasks.

1) *Return donors prediction*: Most of the previous studies applied logistic regression to donor demographic information in order to predict returning donors and to understand the underlying factors affecting this prediction [14][15][16].

A study on the REDS donors' dataset, which had been collected by six blood bank centers around the U.S. from 2003 to 2004, was conducted along with a survey of the donors. The donors' dataset contained information of donation dates, donation status (first-time versus repeat), donation frequency, donation type, and other demographic characteristics (age, sex, and race). The class attribute (1: return donor, 0: non-return donor) was defined as 1 if the donor returned to donate during one year after completing the survey, and 0 otherwise. A binary logistic regression was used to determine the most significant predictive attributes ($p < 0.05$). The study found that predicting donor return was highly dependent on the donation frequency, the convenience of the donation place, and a good donation experience [14].

A study on demographic data, frequency of return, and the time interval between donations was conducted by collecting this data from the Shahrekord Blood Transfusion Center for five years. To conduct this study, researchers created a list of first-time donors for a single year (2008-2009), then the additional variables of frequency of return, the time interval between donations, and the class attribute return to donation (1: return at least once, otherwise 0: non-return) were collected

for the next four years. The three response variables (return to donation, frequency of return and the time interval between donations) were analyzed using logistic regression, negative binomial regression, and Cox's shared frailty model, respectively. The results of the logistic regression showed that donor return was mainly affected by sex, weight, and career [15].

In the U.K., the National Health Service Blood and Transplant (NHSBT) agency conducted a study on a dataset of volunteer donors of whole blood for two years (2010-2011) that included these demographic data: donor ID, postcode, age, ethnicity, donation date, donation type, donor status flag (new or repeat), and most recent previous donation date. The class attribute for return prediction was defined as return donor if the interval between the last donation date and the most recent previous donation date was no more than two years, non-return donor otherwise. Notice that to make sure that first-time donors had enough time to donate again, the research removed all first-time donors that donated in the last six months of the study period. To analyze these features, a multivariate logistic regression was used and showed that donor return varied mainly by geographic location, age, and gender [16].

Other studies relied on survey data rather than blood donation datasets to analyze the factors affecting donors return, with more attention on geographical, educational, and awareness factors [17][18]. Generally, they found that the probability of repeating donation was affected by gender, age, education level, awareness of the donation time, and donor living area.

In addition to logistic regression, other machine learning techniques have been used to predict donor return. For instance, a Classification and Regression Tree (CART) classifier applied to public blood donation data available on the UCI Machine Learning Repository provided a high classification accuracy with 99% precision [19]. In another study with the same dataset, the ANN and SVM algorithms were used for classification with results of ANN sensitivity (65.8%), ANN specificity (78.2%) SVM sensitivity (68.4%), and SVM specificity (70.0%) [20].

2) *Analyzing and forecasting blood donation*: Previous studies in this area of blood donation are few. One similar study on red blood cell (RBC) transfusion by the blood bank of the Hospital Clinic of Barcelona investigated three univariate time-series methods for forecasting the monthly demand by analyzing the RBC demand dataset during a 15-year period (1988-2002). The performance of an autoregressive integrated moving average (ARIMA), a Holt-Winters exponential smoothing model, and a neural-network-based model were compared. The results indicated the excellence the ARIMA and exponential smoothing models for short-time forecasting [21].

The Hacettepe University Hospitals Blood Center conducted a study of investigating donor arrival patterns in order to determine the required workforce for each day of the week and also within a single day. The dataset contained 1095 records, each representing a single day during three years (2005-2007). Each day (record) was described by 20 attributes: year (1-3), month (1-12), day-of-month (1-31), day-of-week

(1-7), and 16 attributes for a one-hour time period of the day (08:00- 24:00) and containing the arrival rate during that hour. The study applied clustering followed by classification methods (rather than a time-series analysis), and considered two donor arrival patterns: daily arrival patterns within the week, and hourly arrival patterns within the day. The results found that the arrival rates to the blood center varied based on ten distinct hourly patterns found within three identified daily patterns (Monday-Thursday, Friday, and Saturday-Sunday) [22].

III. DATA COLLECTION

Data of blood donors had been officially collected from a public Saudi hospital for a period of two years from 2017 to 2018 across various ages, blood types, genders and nationalities. It contains the date of the first donation for each donor and all dates of his/her subsequent donations during the period under study. Table I presents the dataset attribute descriptions.

A. Data Preprocessing

The major steps involved in data preprocessing are feature generation and data cleaning. Feature generation creates new attributes from the original data that can help the modeling. Data cleaning involves detecting outliers and handling noise and missing values.

1) Feature engineering

Three features were added as follows:

- *Last donation date*: This was extracted from the donation date feature by saving all donation dates for a donor during the two years on a sorted list (*DonationDatesList*), then retrieving the last element in this list.
- *Period in months*: This was calculated from the first donation date and last donation date.
- *Class label*: A donor was considered as return donor (1) if he/she donated two or more times within the period under study. Otherwise, the donor was non-return (0). This was calculated by using the length of *DonationDatesList* for each donor. The class is set to 1 if the length is 2 or more, 0 otherwise. Table II shows the total number of instances in each class after adding the attribute to the dataset.

TABLE I. DATASET ATTRIBUTES OF BLOOD DONOR DATA

Attribute name	Attribute type	Explanation
ID	Integer	Donor id
Blood Type	Categorical	The blood type: A+, A-, B+, B-, AB+, AB-, O+, O-
Gender	Categorical	Male (M), Female (F)
Age	Continuous	Range from 18 to 66
Donation count	Discrete	Total donations, from first to last donation dates
First donation date	Date	Date of first donation
donation date	Date	Date of each donation
Nationality	Categorical	Donor's nationality
City	Categorical	Donor's city

TABLE. II. NUMBER OF INSTANCES IN EACH CLASS

Class	Total
0	33018
1	1973
Total	34991

However, donors who donated once and for the first time in 2018 (14000 donors) had been removed from the analysis to ensure that all first-time donors had at least 1 year of follow-up time in which to make a subsequent donation. The reason behind choosing a 1-year follow up period is that the donors who had donated for the first time in 2017 returned in a period mean of 13 months. Among the remaining 21080 donors, 1945 donors were return-donors and 19135 were non-return donors.

2) Data cleaning

- Outliers' detection: There were some outliers in *Age*, *Period in months*, and *Donation count* attributes, as will be discussed later in the analysis section. However, such outliers are valuable to the analysis and will not be removed or changed.
- Missing values: There were missing values in two attributes, which were handled as follows:
 - The nationality variable had one missing value, which was replaced by KSA, as it was the majority value of the variable (99.99%).
 - The city variable had 44% with missing values. However, since Riyadh was the value for 99.99% of the rest, this variable was deleted.

After the preprocessing, the dataset was ready for analysis with the attributes listed in Table III.

TABLE. III. FINAL BLOOD DONOR DATASET

	Attribute name	Attribute type	Explanation
1	ID	Integer	Donor id
2	Blood Type	Categorical	The blood type: A+, A-, B-, AB+, AB-, O+, O-
3	Gender	Categorical	Male (M), Female (F)
4	Age	Continuous	Range from 18 to 66
5	Donation count	Discrete	Number of donations from first to last donation date
6	First donation date	Date	Date of first donation
7	Last donation date.	Date	Date of last donation
8	Nationality	Categorical	Donor nationality
9	Period in months	Discrete	The period in months from first to last donation date
10	Class	Binary	1 if return donor , 0 if non-return donor

IV. METHODOLOGY

This section illustrates the implementation of both classification and time series analysis. For classification, the impact of the imbalance problem in the return donors' prediction is presented by testing the selected classifiers before and after imbalance handling. Classifiers are evaluated in term of performance. Then, the blood donation time series analysis and forecasting are also discussed.

A. Predicting Return Donor

For the classification task that aimed to predict the donors' return, the implementation was conducted in two parts. First, imbalance in the donors' dataset was ignored and three classifiers were built: logistic regression (LG), random forest (RF), and support vector classifier (SVC). Second, the dataset was process while handling the imbalance by over-SMOTE sampling.

In all the constructed models, the following three steps were applied:

- Data Splitting: The data was split into a training set and a testing set. The training set was used to build and validate the models, while the testing set was treated as the new unseen data for evaluation.
- K-fold Cross Validation: A 5-fold cross-validation procedure with the training set was used to train and validate the models.
- Hyper-parameter optimization: A random search approach with cross-validation was used for tuning the hyper-parameters of the models to find the best combination.

The testing results of the three models before handling imbalance problem are shown in Table IV. The low recall of LG and SVC (around 30%) indicated that the models were suffering from an unrecognized positive class, i.e. many return-donors were predicted as non-return donors. On the other hand, in the recall of RF (65%) was higher which demonstrates that RF was better than LG (32%) and SVC (30%) in predicting the minority positive class (return-donor) correctly.

The LG, RF, and SVC models were constructed again, but after handling the imbalance with the Synthetic Minority Over-Sampling Technique (SMOTE). SMOTE was applied to the training set to create synthetic observations of returning donors. The testing results of the three models are shown in Table V.

TABLE. IV. TEST RESULTS FOR ALL CLASSIFIERS WITHOUT HANDLING THE IMBALANCE PROBLEM

Test	LG	RF	SVC
Accuracy	92.92	93.70	92.90
Precision	72.54	64.23	75.11
Recall	32.91	65.83	30.07
F1	45.28	65.02	42.94
AUC	65.85	81.13	64.54

TABLE V. TEST RESULTS WITH RESAMPLING OVER-SMOTE

Test	LG	RF	SVC
Accuracy	92.67	92.64	92.61
Precision	55.14	55.04	54.95
Recall	94.48	94.12	93.77
F1	69.63	69.46	69.29
AUC	93.49	93.31	93.13

The three classifiers show a dramatic increase in recall after handling the imbalance by using over-sampling-SMOTE, from around 30% to 94 % in LG and SVC, and from 65% to 94% in RF. The improved recall demonstrates that the models are better at considering the minority positive class (return-donor) after introducing a balance in the dataset. Moreover, the differences between the LG, RF and SVC are very small for both F1 and AUC, where LG gives the best results.

B. Time Series Analysis and Forecasting

Time series analysis and forecasting from the donors dataset were applied to find any seasonal variations in blood donation that can be identified and communicated to blood banks for better planning and decision-making processes.

1) *Blood donation time series analysis:* The donation date attribute was converted to a date-time object. The dataset was then sorted by date and aggregated on a monthly basis with the *Resample* function. The monthly blood donation is shown in Fig. 2.

From the monthly blood donation plot over the 2-year period, it is possible to note that with the exception of June in each year, the mean shows some consistency. Moreover, the seasonality is pronounced as there are yearly drops in the month of Ramadan (roughly, June, for the period) and to a lesser extent on the Haj month (roughly, September, for the period). These two months match two religious periods (Fasting and Performing Hajj) and also two Islamic Eids, when people in the KSA have celebrate the holidays. During Ramadan, blood donation is not allowed during the fasting time unless as a necessity, e.g. replacement donors.

In order to see the series components, the series was decomposed into trend, seasonality, and residuals as shown in Fig. 3. The time series is clearly stationary, as the level of the series stays roughly constant over time, and the variance of the series appears roughly constant over time.

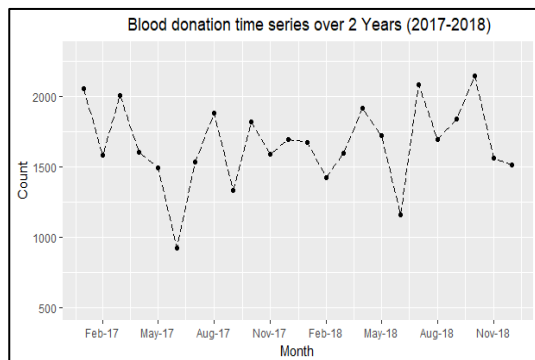


Fig. 2. Blood Donation over 2 Years.

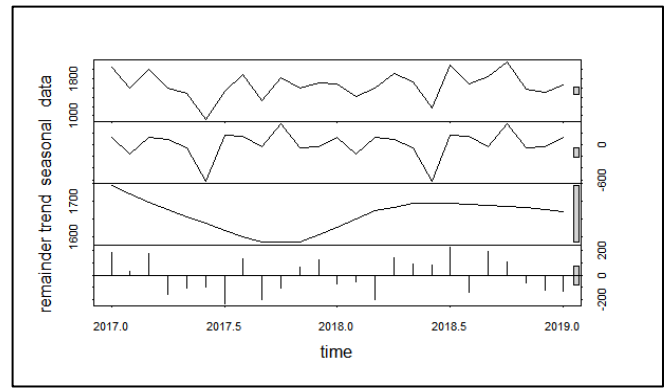


Fig. 3. Blood Donation Time Series Components.

2) *Blood donation time series forecasting using ARIMA:* Since the series appeared stationary, there was no need to apply any differences. The *d* value was set to 0. In order to determine the order of *p* and *q*, the ACF and PACF plots of the original time series were used to check the autocorrelation.

Since there was no significant lag in both ACF and PACF as shown in Fig. 4, the values of *p* and *q* were equal to 0. The final ARIMA model was ARIMA (0, 0, 0), which indicated that the series was a random walk and could not be predicted.

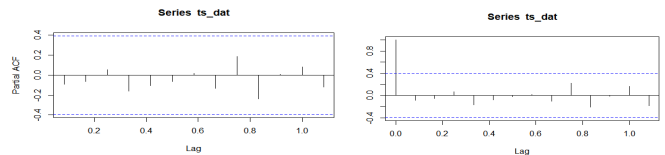


Fig. 4. Auto and Partial Correlation Functions of Blood Donation Time Series.

V. DISCUSSION

The dataset suffers from a limitation in attributes. More attributes can be collected to support the blood center in the prediction task to determine the return donor characteristics, for instance, profession, educational level, weight, neighborhood and reason for donation (volunteer or replacement donor).

As Recommended, hospitals can make use of blood donor's history to offer incentives for returning donors, such as flexible file opening and priority in appointments for them and their families. This would build strong and long-lasting relationships between return donors and the hospitals blood centers. Moreover, blood bank centers should target younger groups, especially in universities. For example, there are more than 40 universities in KSA that average more than 50.000 students each.

VI. CONCLUSION AND FUTURE WORK

In order to predict returning donors, data from donors to a Saudi blood center had been collected over two years (2017-2018). Machine learning algorithms were used to build binary classifiers that were able to predict return-donors with an imbalanced donors' distribution, where few instances belonged to the return-donor class and the majority instances were non-return donors. To illustrate the impact of imbalance problem in

the prediction performance, all classifiers were tested before and after handling the imbalance. Experiments of different classifiers showed very similar performance results.

Moreover, the time series analysis of the monthly donation count explored stable donations over the year with two significant drops occurring during two religion periods, Fasting and Performing Hajj.

For future work, survival analysis and modeling can be applied to predict when the user will return for donation. Moreover, mixed models could be used to find the interaction between variables in the donors' dataset.

REFERENCES

- [1] T. Santhanam and S. Sundaram (2010). "Application of CART algorithm in blood donors classification," *Journal of Computer Science*, vol. 6 no. 5, p. 548.
- [2] B.M. Masser, T.C. Bednall, K.M. White, and D. Terry (2012) "Predicting the retention of first-time donors using an extended theory of planned behavior," *Transfusion*, vol. 52, no. 6, pp. 1303-1310. Available at: 10.1111/j.1537-2995.2011.03479.x.
- [3] "The total population in 2018," General Authority for Statistics. Available at: <https://www.stats.gov.sa/en/indicators/1> [Accessed: February 13, 2019].
- [4] "Ministry of Transport: The number of deaths of traffic accidents in the Kingdom's roads decreased by 33% in 2018," spa.gov.sa. Available at: <https://www.spa.gov.sa/home.php?lang=en>. [Accessed: February 13, 2019].
- [5] M. Alam and D.B.D. Masalmeh (2004). "Knowledge, attitudes and practices regarding blood donation among the Saudi population," *Saudi Medical Journal*, vol. 25 no. 3, pp. 318-321.
- [6] Obeed, M. (2018). "Al-Khashan: 'Health' is moving to encourage citizens to donate blood voluntarily," *Sabq Newspaper Online Edition*. Available at: <https://sabq.org/4FL2xm> [Accessed 3 Oct. 2018].
- [7] O. Filho, W. Cezarino, and G. Salviano (2012). "A decision-making tool for demand forecasting of blood components," *IFAC Proceedings*, vol. 4 no. 6, pp. 1499-1504. Available at: 10.3182/20120523-3-ro-2023.00201.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [8] B. Sojoka and P. Sojka, "The blood donation experience: Self-reported motives and obstacles for donating blood," *Vox Sanguinis*, vol. 94, no.1, pp. 56-63. Available at: 10.1111/j.1423-0410.2007.00990.x.
- [9] C. Holly, S. Balegh, and B. Ditto (2011). "Applied tension and blood donation symptoms: The importance of anxiety reduction." *Health Psychology*, vol. 30 no. 3, pp. 320-325. Available at: 10.1037/a0022998.
- [10] M. Anwer, S. Ul Fawwad, S. Anwer and A. Ali (2018), "Attitude toward blood donation among medical and nonmedical students across Karachi," *Asian Journal of Transfusion Science*, vol. 10, no. 2, p. 113. Available at: 10.4103/0973-6247.187937.
- [11] A. Abdel Gader, F. Al Gahtani, A. Ramadan, A. Osman, M. Farghali and A. Al-Momen (2011) "Attitude to blood donation in Saudi Arabia," *Asian Journal of Transfusion Science*, vol. 5 no. 2, p. 121. Available at: 10.4103/0973-6247.83235.
- [12] S. Hadi Alharbi et al. (2018). "Assessment of levels of awareness towards blood donation in Saudi Arabia," *AIMS Public Health*, vol. 5 no. 3, pp. 324-337. Available at: 10.3934/publichealth.2018.3.324.
- [13] S. Elsafi, M. Al Zahrani, and E. Al Zahrani (2015). "Awareness and practice of blood donation by college students in Dhahran, Saudi Arabia," *ISBT Science Series*, vol. 10 no. 1, pp. 11-17. Available at: 10.1111/voxs.12172.
- [14] K.S. Schlumpf, et al. (2008). "Factors influencing donor return," *Transfusion*, vol. 48 no. 2, pp. 264-272.
- [15] S. Kheiri and Z. Alibeigi (2015). "An analysis of first-time blood donors return behaviour using regression models," *Transfusion Medicine*, vol. 25 no. 4, pp. 243-248.
- [16] S. Lattimore, C. Wickenden, and S. Brailsford (2014). "Blood donors in England and North Wales: Demography and patterns of donation," *Transfusion*, vol. 55 no. 1, pp. 91-99. Available at: 10.1111/trf.12835.
- [17] T. Volken, C. Weidmann, T. Bart, Y. Fischer, H. Klüter, and P. Rüesch (2013). "Individual characteristics associated with blood donation: A cross-national comparison of the German and Swiss population between 1994 and 2010," *Transfusion Medicine and Hemotherapy*, vol. 40 no. 2, pp. 133-138.
- [18] W.I. Mauka, M.J. Mahande, S.E. Msuya, and R.N. Philemon (2015). Factors associated with repeat blood donation at the Northern Zone Blood Transfusion Centre in Tanzania," *Journal of Blood Transfusion*, vol. 2015.
- [19] T. Santhanam and S. Sundaram (2010). "Application of CART Algorithm in Blood Donors Classification," *Journal of Computer Science*, vol. 6 no. 5, pp. 548-552. Available at: 10.3844/jcssp.2010.548.552.
- [20] M. Darwiche, M. Feuilloley, G. Bousaleh, & D. Schang, D, "Prediction of blood transfusion donation," presented at Research Challenges in Information Science (RCIS), 2010 Fourth International Conference (2010, May).
- [21] A. Pereira (2004). "Performance of time-series methods in forecasting the demand for red blood cell transfusion," *Transfusion*, vol. 44 no. 5, pp. 739-746. Available at: 10.1111/j.1537-2995.2004.03363.x.
- [22] M. Testik, B. Ozkaya, S. Aksu and O. Ozcebe, "Discovering Blood Donor Arrival Patterns Using Data Mining: A Method to Investigate Service Quality at Blood Centers," *Journal of Medical Systems*, vol. 36 no. 2, pp. 579-594. Available at: 10.1007/s10916-010-9519-7.