iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.41940

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Forecasting System for Donation of Blood Using SVM Model

Pooja Selvaraj[1], Aiman Sarin[2], B. Ida Seraphim[3]
[1, 2, 3]*Department of Computer Science and Engineering, SRM Institute of Science and Technology*

Abstract: ***The act of donating blood allows an individual to become a responsible citizen of the society by giving life to others by increasing the effectiveness and efficiency of a great practice. Forecasting blood supply is a serious and recurrent problem for blood collection managers. Machine Learning can be used to learn the patterns in the data to help to predict future blood donations in turn saving more lives. This will establish an efficient link between two major aspects of the blood donation process i.e. blood donation centres where blood is donated and hospitals where blood is in regular demand.***
Keywords*: Blood Donation Prediction System, ML Techniques, Web App, Support Vector Machines*

I. INTRODUCTION

The primary concern, as per the results of the cross-sectional study, seems to be the delay in identifying the blood donation facilities since individuals are uninformed of surrounding blood donation centres. This is due to the lack of effective relationships linking clinics and blood donation centre that can ensure that plasma is accessible to patients when they need it. Blood donation clinics, however, lack the resources necessary to identify potential donors.

Blood transfusion helps to save a lot of lives but at the same time having access to a healthy blood group free of illness in dire emergency time period is the real challenge that bridges the gap between life and death. The activities related to blood collection, blood screening, testing, processing, storage, distribution have to be effectively coordinated and organized in a sophisticated manner with the help of integrated blood supply networks.

The primary goal of this study is to address this issue by establishing a well-organized link amongst the blood donation centres and clinics in order to better forecast blood donations from probable donors. This will help patients' lives and will make it simpler for them by allowing them to obtain blood from local locations when they are in need. We also want to raise awareness about blood donation and urge others to practise the same. We'll be using information from the Blood Transfusion Service Centre's donor record in Hsin-Chu City, Taiwan [1]. The facility sends a blood transfusion service bus to a campus in Hsin-Chu City every 3 months to collect blood donations. We will create a ML model that can forecast whether or not a blood donor will donate again.

II. DATABASE

The Blood Transfusion Service Dataset contains five primary variables shown in Figure 1. Each attribute has a numeric datatype, which is ideal for building an artificial intelligence model. Table I depicts a portion of the data file that we will be using [2]. The expert system model that is being built is a binary classifier, which provides us with just two outcomes:

1) Donor will not donate with result as '0'
2) Donor will donate blood with result as '1'

1. Recency: no. of months since last donation
2. Frequency: total no. of donations
3. Monetary: total volume of blood donated in cubic centimetres
4. Time: no. of months since first donation
5. A binary variable indicating if he/she donated blood in March 2007 – where 1 indicates that he/she donated; 0 indicates that he/she did not donate

Fig. 1 Five primary variables

TABLE I
SNIPPET OF UCI DATASET

Recency	Frequency	Monetary (cm3)	Time	Binary variable
2	4	1000	14	1
0	2	500	4	0
4	7	1750	25	1
1	9	2250	51	0
2	4	1000	16	1

III. LITERATURE SURVEY

A literature study of past work in this topic, i.e. blood donation predictive model employing diverse methodologies, was undertaken using various academic sources. Three research articles were considered in order to conduct suitable analysis in the area of our interest.

A. Blood Donation Interval Estimation through Deep Learning

According to P. KIRCI, blood transfusion is critical for human survival in life-threatening situations. As a result, a range of machine learning methods can be used to estimate the number of potential donors and the likelihood of blood donation [3].
By leveraging datasets, machine learning algorithms assist blood transfusion procedures. Machine learning is a technique for finding patterns in massive volumes of data. Algorithms use the data provided to complete learning and machine tasks. On the blood transfusion data set, one of the most successful classification algorithms was compared based on performance.
Several countries use various tactics to increase the number of blood donors, such as South Africa's Club 25 and Ghana's national blood policy, which govern the blood donation and transfusion process while also attempting to educate, encourage, attract, and retain young blood donors.

B. Forecasting Blood Donor Response using Predictive Model Approach

In this study, we look at whether we can predict whether or not a person would donate blood using previously untested binary classification algorithms on grouped and non-clustered data [4].
Blood banks may now use their large database to better prepare for future crises while also saving lives, thanks to Big Data's dramatic growth in the amount of data available. Many people die each year owing to a scarcity of blood and their inability to receive it in a timely way, therefore this research strives to make the greatest use of blood donated by donors.

C. Predicting the behaviour of Blood Donors in National Bank of Ethiopia Using Data Mining Techniques

In this paper, they have used data mining techniques for implementation in health centres to forecast the behaviour in blood donors. They have used techniques like J48 decision trees, neural networks, and Naive Bayes [5]. Since human data analysis is insufficient when dealing with massive amounts of data, computer based data mining technology for data analysis is sufficient. This method will be used to search the Harar branch health centre's database of blood group donors. This algorithm will be tweaked to determine the conditions under which blood donors are properly qualified to give in an emergency and assist in the collection of safe blood.

IV. METHODOLOGY

The Cross-Validation is a technique for determining the efficacy of various machine learning models when they are run on a testing/validating dataset. For our dataset, Logistic Regression Classifiers, Support Vector Machines, and Random Forest Algorithm are the most successful algorithms [6]. Since the training as well as cross-validation curves are further together after grid search optimization in Figure 2 and Figure 3, we may deduce that the SVM Classifier and Random Forest classifiers appear to generalize the prediction better.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



Fig. 2 Learning Curve for SVC

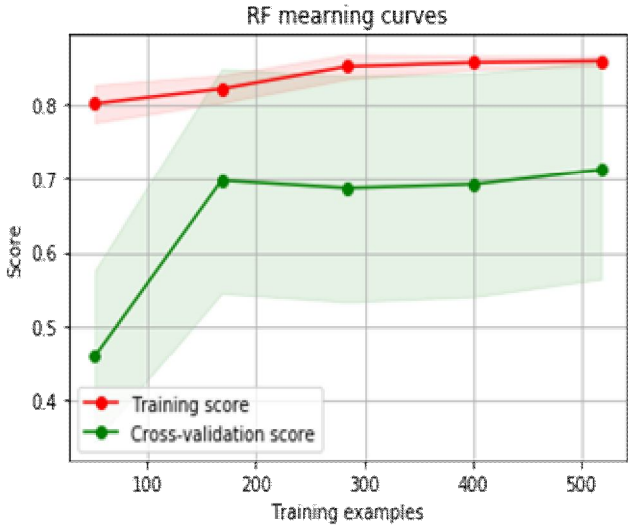


Fig. 3 Learning Curve for Random Forest

A. Exploratory Data Analysis

In layman's terms, it is simply a method of analyzing data through the use of visual approaches with the use of statistical summaries and graphical representations. This allows us to find trends, patterns, and to understand the various aspects of the data. In order to perform the necessary steps for exploratory data analysis, we first import the necessary libraries along with the train and test dataset. The basic and systematic steps that we have followed in order to explore the data in the most efficient way possible is as follows:

1) *Understand the data:* In this step, we can use various functions to identify the number of rows and columns, to infer some important features from the column names, and to comprehend the statistical data. We found that there were no missing values neither any outliers in our data. Our data is a bit under sampled for the target variable which could be improved later.
2) *Clean the data:* Here, we should clean our data from redundancies as can be irregulatires or columns that are unnecessary so we can remove them. Incomplete, inaccurate, or noisy data, as well as duplicated information, can be found in raw data. In this step, we can remove highly correlated features and/or add extra variables. In our case, we have added a new variable 'Donating for' which is simply 'Months since first donation' subtracted by 'Months since last donation'.
3) *Analyze the relationship between the variables:* As there are multiple features in a dataset, few of them can be correlated to each other. While training our model, these correlated features may cause our model to be biased, hence these should be dropped from the dataset before training our model.

a) *Correlation matrix:* This feature allows us to define the relation among two variables. Hence, the correlation matrix is a table that depicts association between the variables as shown in Figure 4.

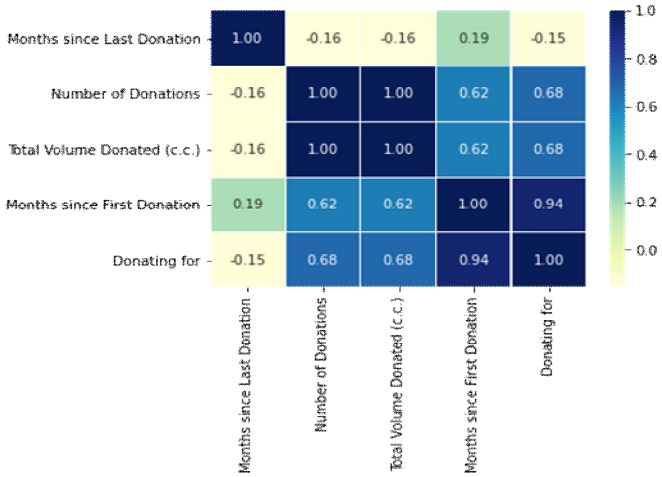


Fig. 4 Correlation Matrix.

B. Training the Model

As we observed from the cross-validation models, both SVC and Random Forest performed well for the given dataset. Although due to numerous benefits, Support Vector Machines is the most widely employed algorithm in the field of healthcare. As a result, it is critical to grasp this approach, which will be quite valuable when applied to the healthcare industry [7].

1) *Advantages:* SVM is beneficial for applications with small sample size which is perfect for our dataset as we are working with less than 800 rows only. Random forest would have been more beneficial when dealing with large amount of data, which would have allowed it to give better accuracy as number of decision trees would have also been higher.
2) *Accuracy:* Our machine learning model, which uses a Support Vector Classifier, has an accuracy of roughly 78.4 percent. It's merely a basic approach to assess the model's efficacy. Although, accuracy is not a suitable parameter to know how excellent the model is in an under sampled database, this must not be the primary estimation of how fit the prototype is. The AUC created by the ROC curve is a recommended statistic for determining which model performs best among the others. Figure 5 shows the Receiver Operating Characteristic curve for our ML model.

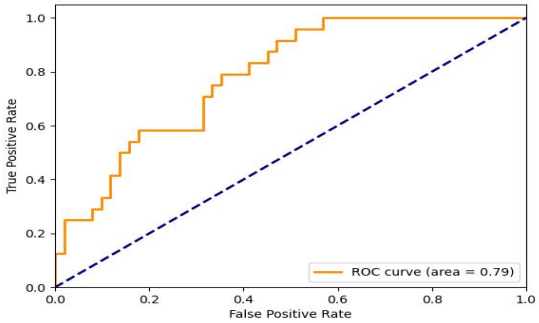


Fig. 5 Final ROC Curve on the test set.

V. INFERENCES

An algorithm to determine whether or not a person would donate again was developed. This algorithm was used in the development of a blood donation forecasting system. The anticipated output can be either a '1' indicating that the donor will contribute again or a '0' indicating that the donor is unlikely to donate again.

Only 748 contributors were included in the dataset we gathered. To further assist the prediction model, more tuples of data might be collected. By adding more data, the accuracy of the results may improve. Other machine learning techniques, in addition to Support Vector Machines, can be employed to provide solid results.

By gathering blood donor data from numerous blood banks in respective cities, hospitals may make better use of the forecasting system. This is essential because you never really know when someone may require a blood transfusion.

VI. CONCLUSION

Advancing in the area of blood donation predictions is the motive of this study. The current machine learning model was constructed utilizing the support vector classifier approach with four components and has an accuracy of around 79.1 percent (recency, frequency, monetary, time).

REFERENCES

[1] Pooja Selvaraj, and Aiman Sarin. "Blood Donation Prediction System Using Machine Learning Techniques." Blood Donation Prediction System Using Machine Learning Techniques, IEEE Xplore, 31 Mar. 2022, https://ieeexplore.ieee.org/document/9740878/keywords#keywords.

[2] UCI Machine Learning Repository: Blood Transfusion Service Center Data Set. https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center. Accessed 22 Nov. 2021.

[3] Borhade, Komal Laxman, et al. "Blood Donation Interval Estimation through Deep Learning." JETIR, JETIR(Www.jetir.org), June 2021, https://www.jetir.org/view?paper=JETIR2106294.

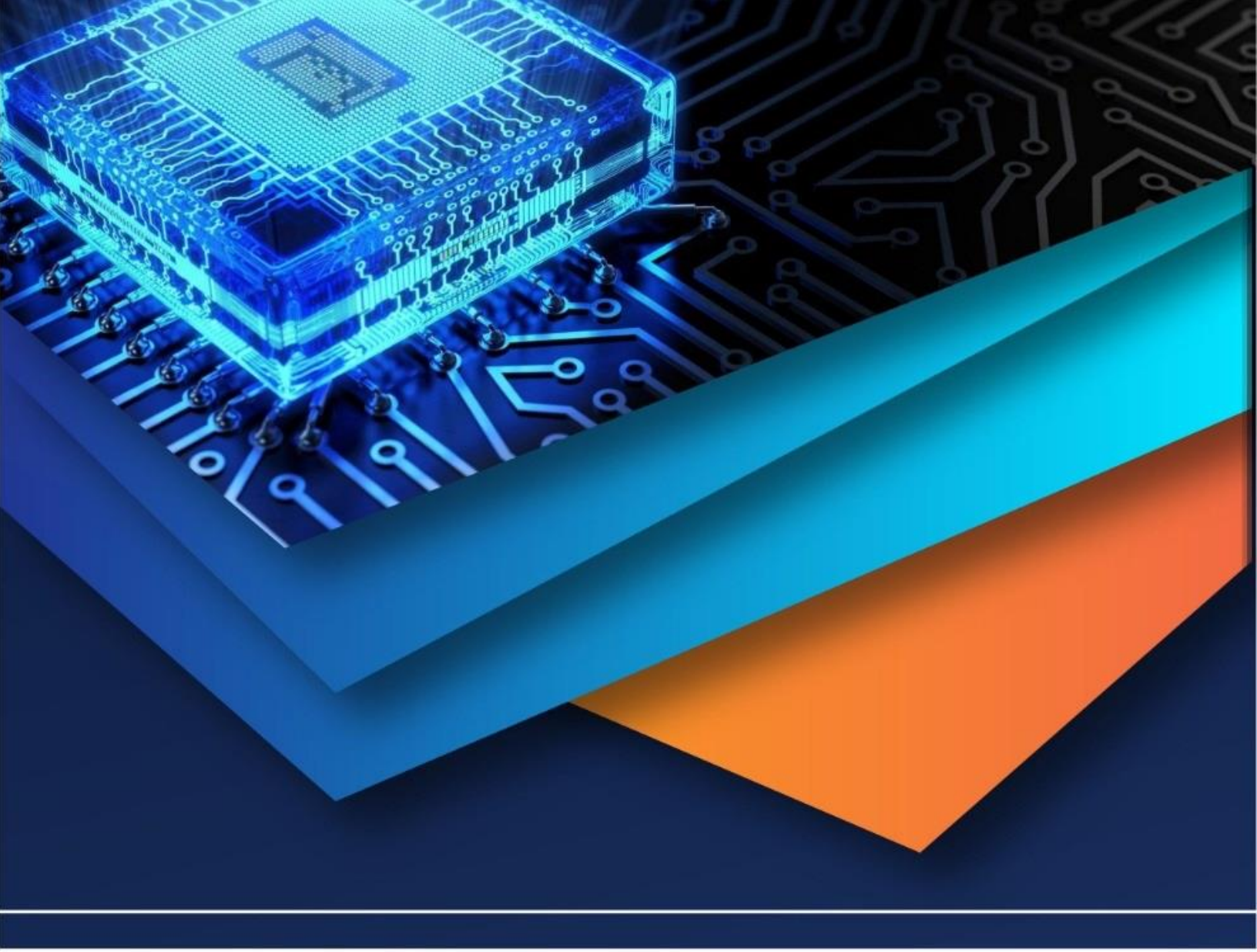
[4] Pradeep, Aalekh, et al. "Forecasting Blood Donor Response Using Predictive Modelling Approach." Forecasting Blood Donor ResponseUsing Predictive Modelling Approach, IJCSMC, 25 Apr. 2019, https://www.academia.edu/38815013/Forecasting_Blood_Donor_Response_Using_Predictive_Modelling_Approach_.

[5] Birhane, Teklay, and Brhanu Hailu. Predicting the Behavior of Blood Donors in National Blood Bank of Ethiopia Using Data Mining Techniques, MECS, June 2021, https://mecs-press.org/ijieeb/ijieeb-v13-n3/IJIEEB-V13-N3-5.pdf.

[6] S. Bhange, "Predicting If a Blood Donor Will Donate within a given Time Window..." Medium, Jovian-Data Science and Machine Learning, 30 Sept. 2020, https://blog.jovian.ai/predicting-if-a-blood-donor-will-donate-within-a-given-time-window-8dbf41a4fe97.

[7] S. Shah Huma. "5 Data Science Projects in Healthcare That Will Get You Hired." Medium, Towards Data Science, 25 Oct. 2021, https://towardsdatascience.com/5-data-science-projects-in-healthcare-that-will-get-you-hired-81003cadf2f3.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)