



CSAC224-MachineLearning

Blood Transfusion Service Center Prediction Using Machine Learning

Jesrel Pizzaro

College of Computer Studies
CamarinesSur Polytechnic Colleges
jepizzaro@my.cspc.edu.ph

Abstract

The motivation of this study is to describe to face the challenge of solving unsolved the problem and to get the intellectual joy of doing this paper. The importance of transfusion of blood will also be one way to save a person's life, so it essential to monitor the blood availability based on its type to avoid an inadequate supply of blood. My objective in this study is measure the level of knowledge regarding the blood transfusion. In the methodology it is predictive modeling Techniques investigated the study a Random forest, XG Boost, a Voting Classifier, and individual ML models logistic regression, KNN, decision tree, and gradient boosting. The model's performance is evaluated using evaluation measures such confusion matrix, precision, recall, F1-score, accuracy, and classification report after the dataset has been preprocessed by scaling features. In order to integrate the advantages of several models, ensemble techniques like Voting classifier and Stacking Classifier are also used. With the Stacking Classifier obtaining accuracy is 80% on the test set. The results show excellent performance and demonstrate the value of integrating various ML techniques.

1 Introduction

The critical issue in Healthcare by predicting an individual is likely to donate the blood. The blood donation plays a crucial role in ensuring a stable and sufficient blood supply for medical treatments, emergencies, and various health conditions. Predicting donor behavior can aid blood centers in developing targeted strategies to encourage and optimize blood donation campaigns. The problem statement revolves around the classification of the individuals based their decency, frequency, monetary contribution, and the time since their last blood donation on March 2007. The binary output represents whether an individual donated blood (1 for donation, 0 for no donation). This classification task is vital for blood centers to identify potential donors and tailor outreach efforts effectively. My motivation for this project stems from significance of maintaining an ample and consistent blood supply, which is crucial for saving lives and providing timely medical interventions. Blood donation patterns are influenced by various factors, and predicting resources efficiently, minimizing shortages, and planning targeted awareness campaigns. The blood donation patterns are influenced by diverse factors, including personal preferences, awareness, and societal consideration. Understanding and predicting these patterns can helps blood centers create proactive strategies to engage donors and ensure a sustainable blood supply. Machine learning algorithms and ensemble methods provide a promising approach to analyze and model these complex relationships. The input to our algorithms consists of historical data related to an individual's blood donation behavior, while the output is a binary prediction indicating the likelihood of the person donating blood. This explicitly definition of input and output parameters ensures clarity and facilitates understanding across different teams and application domains

2 Related work

I gathered some related studies based on given datasets. The several studies have explored the use of machine learning algorithms and ensemble methods in the field of blood transfusion classification. For example, the paper Machine Learning for Blood Donors Classification Model shows how to apply machine learning classification algorithms to locate homogeneous groups of blood donors and forecast factors leading to donor return. This is done by employing ensemble methods. Predicting Blood Donors Using Machine Learning Techniques is another study that addressed problems in the US blood supply chain brought on by advancements in hospital policy, transfusion management, and surgical practice. It did this by analyzing data from regional blood centers. And last, the ongoing need for blood transfusions, which are necessary for a number of medical procedures and life-or-death surgeries, served as the driving force behind the development of Machine Learning for Blood Donors Classification Model using Ensemble Learning. And additionally, in the review machine learning in transfusion medicine. In a scoping review, the major methodological methods and current trends in applying machine learning to transfusion medicine are discussed, along with the difficulties and potential solutions for its future application. Finally, the application of contemporary machine learning techniques to create and evaluate a model that can precisely forecast the requirement for a huge blood transfusion was investigated in the paper Assessment of Machine Learning Methods to predict huge Blood. These studies have made significant strides in using machine learning techniques, including ensemble learning, to predict blood donation patterns and needs. However, there is still room for improvement, particularly in terms of accuracy and real-time prediction capabilities. The integration of more advanced technologies, such as deep learning, could potentially enhance these prediction models.

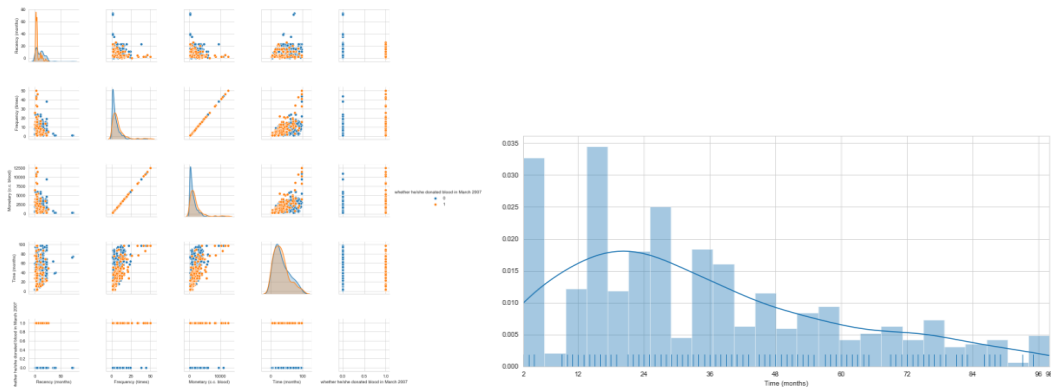
CS230: Deep Learning Winter 2007, Stanford University, CA. (LateXtemplateborrowedfromNIPS2017.

3 Data sets and Features

In training the model I used straightforward to function, which model fit. In the algorithm I used 50% of the training data for validation and set the test size 25%. I split the data, and set the holdout variables in 75% the length of the data for training and testing. Before test of the model, I print the content of the data frame and normalize the data to check the value. The time series of the data, it show the high range of “regency”, ”frequency”, ”monetary”, and the people who donated blood. I also use the histogram of the plot the ranges of each data and include a seaborn plot to see the input is correlated to each other based on the class.

<http://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>

	Regency (months)	Frequency (times)	Monetary (c.c. blood)	Time (months)	whether he/she donated blood in March 2007
count	748.000000	748.000000	748.000000	748.000000	748.000000
mean	9.506684	5.514706	1378.676471	34.282086	0.237968
std	8.095396	5.839307	1459.826781	24.376714	0.426124
min	0.000000	1.000000	250.000000	2.000000	0.000000
25%	2.750000	2.000000	500.000000	16.000000	0.000000
50%	7.000000	4.000000	1000.000000	28.000000	0.000000
75%	14.000000	7.000000	1750.000000	50.000000	0.000000
max	74.000000	50.000000	12500.000000	98.000000	1.000000



4 Methods

My machine learning algorithms and ensemble methods for the classification task related blood transfusion. The primary algorithms employed include random forest, XG Boost, a Voting Classifier combining Logistic Regression, decision Tree, SVM, KNN and Gradient Boosting, as well as individual models of Logistics regression. The datasets containing information such as decency, frequency, monetary contributions, and time since the first donation is meticulously processed and examined for missing values. Descriptive statistics and correlation heat maps provide insight into dataset's characteristics. The data preprocessing involves scaling the features, enhancing the models' performance. Subsequently, the dataset is split into training and testing sets. Each algorithm is trained on the training set and evaluated on the test set using diverse evaluation metrics such as confusion matrices, precision, recall, F1-score, and accuracy. Additionally, a voting classifier and stacking classifier are employed to leverage the strengths of multiple models. The voting classifier amalgamates combines the predictive capabilities of logistic regressions, Decision tree, SVM, KNN, and Gradient Boosting in a meta- classifier framework. The code conclude with a comparison of the performance of these models, aiding in the identification of the most effective approach for predicting blood donation behavior based on the provided dataset. Overall, the code provides a comprehensive analysis and implementation of various machine learning techniques for the specific blood transfusion classification problem.

Here is the equation provided:

$$j(w^t, b) = \frac{1}{m} \sum_{i=1}^m L(y^1, y^2)$$

5 Experiment/Result/Discussion

The exploratory data analysis (EDA) of the dataset reveals intricate relationships and patterns, such as the direct correlation between the frequency of blood donations and their monetary value. The analysis uses a wide range of machine learning models, including decision trees, logistic regression, random forests, K-Nearest Neighbors, gradient boosting, and support vector machines. With impressive 80% accuracy, the support vector classifier with scaled features stands out as the most promising model. The code demonstrates a thorough method for selecting and evaluating models by methodically using hyperparameter tuning and investigating an ensemble model. As we go, a number of directions for future research become apparent. First, more feature engineering and extraction may be investigated in order to improve the models' capacity for prediction. This could entail drawing new features from the already-existing ones or integrating data from outside sources. Further examination of the temporal facets of blood donation, such as seasonality or long-term trends, may also yield insightful results. Additionally, doing a more thorough hyperparameter tuning grid and investigating cutting-edge optimization strategies might enhance the support vector classifier's performance even more. Lastly, the addition of explainability tools like LIME or SHAP values may help to improve interpretability by illuminating the model's decision-making process.

Primary metrics, we will use the following formula:

Accuracy Score = (TP+TN)/ (TP+FN+TN+FP)

Recall Score = TP/ (FN+TP)

Precision Score = TP/ (FP+TP)

F1 Score = 2* Precision score* Recall Score/ (Precision Score + Recall score/)

In the perceptron model the accuracy score is 78%, and its across validation 76% accuracy, here is the classification report.

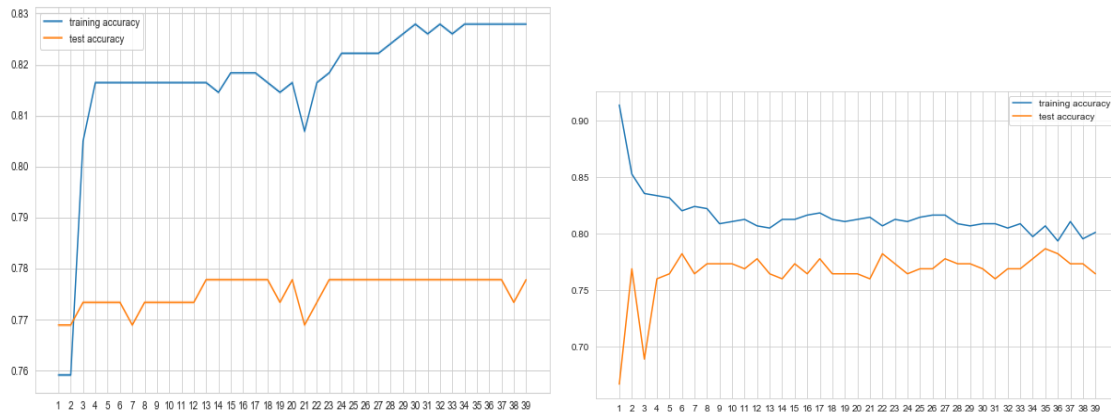
	Precision	recall	f1-score	support
0	0.82	0.92	0.87	173
1	0.55	0.31	0.40	52
Avg/total	0.76	0.78	0.76	225

In K Neighbors Classifier model the accuracy score is 77 %, and its cross-validation score is 79%,

Here is the classification report of K Neighbors Classifier model

	Precision	recall	f1-score	support
0	0.81	0.97	0.88	173
1	0.72	0.25	0.37	52
avg / total	0.79	0.80	0.77	225

In the Neural network, its accuracy score is 82.98%, and its test loss is 28%, here is accuracy, I improvement with epoch: in the shown plot, it fluctuates the accuracy value of the training dataset.



6 Conclusion/Future Work

In conclusion, the algorithm that has the highest performing rate is the neural network. There are some algorithms that worked better than others depending on the nature of the data that we use, and what pre step and method we implemented. Each dataset has the best match machine learning algorithm. If we had more time, more members, or more computational resources, we would like to explore further the different and latest machine learning algorithms, to find out what the best algorithm will work given the different datasets.

7 Contribution

I'll responsible the introduction, conclusion, and rendering the data. And related works and rendering the data. The abstract, methodology and rendering the data. I'll make sure i finalized my work.

8 References

- [1] Sander A, Insisters J, Gumboots H. Patient blood management: the global view. Transfusion. 2016 (supply 1):S94–S102.
- [2] Ellington KD, Sapient MRP, Harass KA, et al. Continued decline in blood collection and transfusion in the United States-2015. Transfusion. 2017; 57(supply 2):1588–1598. [PMC free article]
- [3] Leahy MF, Hofmann A, Towner S, et al. Improve outcomes and reduced costs associated with a health-system-wide patient blood management program: a retrospective observational study in four major adult tertiary-care hospitals. Transfusion. 2017; 57 (6):1347–1358.
- [4] Hofmann A, Farmer S, Towner SC. Strategies to preempt and reduce the use of blood products: an Australian perspective. Corroding Anesthesia. 2012; 25 (1):66–73.
- [5] Sander AS, Good enough LT. Blood transfusion as a quality indicator in cardiac surgery. JAMA. 2010; 304 (14):1610–1611.
- [6] Shih H, Reentrant S. Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation blood supply. J Health Eng. 2019; 2019:1–6. PMC free article.
- [7] Gumboots H, Rehab PH, Sander A, Hofmann A. Blood use in elective surgery: the Austrian benchmark study. Transfusion. 2007; 47(8):1468–1480.

[8] Jalali, Ali PhD*; Lonsdale, Hannah MBChB†; Zamora, Lillian V. MD†; Ahumada, Luis MSCS, PhD*; Nguyen, Anh Thy H. MSPH‡; Rehman, Mohamed MD†; Fackler, James MD§; Stricker, Paul A. MD‖; Fernandez, Allison M. MD, MBA†; Pediatric Craniofacial Collaborative Group. Machine Learning Applied to Registry Data: Development of a Patient-Specific Prediction Model for Blood Transfusion Requirements During Craniofacial Surgery Using the Pediatric Craniofacial Perioperative Registry Dataset. *Anesthesia & Analgesia* 132(1):p 160-171, January 2021. | DOI: 10.1213/ANE.0000000000004988

[9] Liu, Le-Ping and Zhao, Qin-Yu and Wu, Jiang and Luo, Yan-Wei and Dong, Hang and Chen, Zi-Wei and

[10] Gui, Rong and Wang, Yong-Jun. Machine Learning for the Prediction of Red Blood Cell Transfusion in Patients During or After Liver Transplantation Surgery. *Frontiers in Medicine* 2021; doi:10.3389/fmed.2021.632210. <https://www.frontiersin.org/articles/10.3389/fmed.2021.632210>

