Information Systems Frontiers
https://doi.org/10.1007/s10796-021-10149-1

Predicting Blood Donors Using Machine Learning Techniques

Christian Kauten[1] · Ashish Gupta[2] · Xiao Qin[1] · Glenn Richey[3]

Accepted: 20 May 2021

Abstract
The United States' blood supply chain is experiencing market decline due to recent innovations in surgical practice, transfusion management, and hospital policy. These innovations strain US blood centers, resulting in cuts to surge capacities, consolidation, and reduced funding for research and outreach programs. In this study, we use data from a regional blood center to explore the application of contemporary machine learning algorithms for modeling donor retention. Such predictive models of donor retention can be used to design more cost effective donor outreach programs. Using data from a large US blood center paired with random forest classifiers, we are able to build a model of donor retention with a Mathews correlation of coefficient of 0.851.

Keywords Analytics · Blood donors · Blood supply · Machine learning · Retention

1 Introduction

Unlike other markets, blood collection has a unique structure with a purely stochastic input. Few hospitals collect blood internally and instead rely on independent blood centers for collection services (Ellingson et al., 2017). Blood centers coordinate with hospital blood banks to provide sufficient capacity for transfusion services. Although some contractual obligations are in play, centers regularly cold-call hospitals daily to determine hospital needs (Mulcahy & Health, 2016). Oddly, hospitals do not inherit financial liability for units of blood until transfusion; blood centers are monetarily accountable for expired units. As such, centers work to reduce waste by swapping out old units with fresh ones, routing the old units to hospitals with dense emergency needs (Beliën & Forcé, 2012). These unusual conditions give hospitals buying power in the market and place strain on blood centers (Mulcahy & Health, 2016). Human donors attend mobile blood clinics hosted at local churches, schools, and other social areas to give whole blood or targeted components like red blood cells, platelets, or plasma through a process called apheresis (Mulcahy & Health, 2016; Pierskalla, 2005). Blood centers typically model this stochastic input statistically, using previous years' usage to predict the supply for the current year. These mathematical rules are brittle, generalizing poorly to a market undergoing constant change. This consideration draws attention to the need for a general model robust against fluctuations in the market.

A recent recession in blood usage places strain on blood centers. In 2011, The National Blood Collection and Utilization Survey (NBCUS) describes the first appearance of decline with significant reductions in both collection and transfusion of Red Blood Cells (RBCs). The report lists Patient Blood Management (PBM) programs, preoperative interventions, and strict transfusion guidelines as potential causes for these decreases (Whitaker et al., 2013). The authors of the 2015 report cite technological innovations in surgery, recent findings in transfusion research, and more systematic management of transfusions as potential

✉ Ashish Gupta
azg0074@auburn.edu

Christian Kauten
kautenja@auburn.edu

Xiao Qin
xqin@auburn.edu

Glenn Richey
richey@auburn.edu

[1] Computer Science and Software Engineering, Samuel Ginn College of Engineering, Auburn University, Auburn, AL 36849, USA

[2] Department of Systems & Technology, Harbert College of Business, Auburn University, Auburn, AL 36849, USA

[3] Department of Supply Chain Management, Harbert College of Business, Auburn University, Auburn, AL 36849, USA

Published online: 17 July 2021

causes for this decline (Ellingson et al., 2017). Hospitals maintain strong buying power over blood centers and dictate market price despite these decreases in demand and supply (Mulcahy & Health, 2016). These considerations place stress on blood centers, resulting in surge capacity reductions, consolidation, and cuts to research and outreach programs.

Outreach programs include a variety of solicitation methods applied by blood centers to bring new or existing donors in to give blood. These include campaigns over mail, e-mail, phone, and other communication platforms (Miah, 2020). Because blood supply is purely altruistic, these programs are necessary to probe the community for donations. However, when considering registered donors, a high rate of donor attrition results in a significant amount of fruitless outreach that could be pruned to reduce costs without heavily impacting the efficacy of such programs. In order to reduce the quantity of e-mails, phone calls, mail flyers, etc., predictive models that identify donors that are likely to return (i.e., will not defer) can be employed.

In this study, we use operational data from a large regional United States (US) blood center to develop a machine learning model for predicting blood donors. We show that such a model allows the blood center to forecast returning donors with high confidence. Our study of machine learning is motivated by the observation that public sectors, such as the blood industry, have not adapted to the era of big data despite the observable positive impact of big data and analytics in other areas of medicine, both public and private (Klievink et al., 2017; Gupta et al., 2018).

Our study addresses the following research questions:

1. Which machine learning methods are effective for predicting returning blood donors in a time window of one month?
2. What hyperparameter configurations produce the most effective models in terms of testing metrics?
3. Which features of the engineered dataset provide the most salient contribution to the predictions?

A large US blood center, namely, Blood Center (Name blinded) (BC), provides a full copy of its operational database from 2015–2017 to analyze in this work. We develop machine learning–based donor retention models which would allow BC recruitment and marketing teams to design better outreach programs. Starting with a large operational database allows for the development of a dataset with more features, higher variance, and less bias than existing blood donor dastasets (i.e., the UCI Blood Transfusion Service Center dataset). The primary contributions of this study are: (1) a processing pipeline that transforms relational data from BC databases into a tabular donor classification dataset for machine learning algorithms, (2) machine learning models capable of predicting returning donors with 80% sensitivity and 99% specificity for a regional US blood center; and (3) detailed discussion of the variables determining donor prediction based on feature importance's derived from trained Random Forest and Gradient Boosting models.

We organize the remainder of this paper as follows. In Section 2, we discuss related literature and draw comparisons to this work. Next, section 3 describes the methodology for modeling donor retention in the BC data. We go on to present the results of the donor retention models in Section 4. We discuss the implication and limitations of this work in Section 5 and provide conclusions and future research directions in Section 6.

2 Related Work

Osorio et al. (2015) posed a comprehensive review of models in blood supply chain management organized into four sequential stages of the process: collection, production, inventory, and distribution. Within the collection stage, sub-problems emerge – i.e., event and staff scheduling, donor recruitment and retention, equipment management, and blood extraction and transportation. Because disasters can adversely affect the ability of the existing donor pool and qualified staff as well as massively disrupt transportation routes, the collection stage is of particular interest. Furthermore, as the entry point of the sequential process, disruptions in the collection phase affect downstream stages. This section addresses contemporary works that solve different problems in the collection phase of blood supply chain management.

2.1 Shelf Life

Abbasi and Hosseinifard (2014) discuss the issuing policy of a perishable inventory, namely RBC, and apply a modified FIFO policy to reduce the age of issue of units of RBC. Abbasi et al. (2017) use a discrete event simulation model to show that reducing the shelf life of RBC for a blood center in New South Wales results in fresher blood transfusions for patients, but imposes more significant risks due to potential shortages. A potential reduction in shelf life of RBC would imply an increase in the number of required donations. To reduce waste, shelf allocation policies, such as the one posed by Kamyabniya et al. (2018), can inform blood centers as to an optimal allocation of blood products among various centers, hospitals, etc.

2.2 Donor Intentions and Motivations

Masser et al. (2011) investigated donor intentions during ongoing epidemics where vaccines are unavailable (e.g.,

swine flu (H1N1)). Although the decision to donate is known to be a rational process, there are outstanding factors that affect an individual's motivations. These factors include, but not limited to, fear of sickness, desire for altruism, and the like. Using a psychological model of the decision process and survey instruments, Masser et al. (2011) discover that individuals who (1) perceive less danger of viral transmission and (2) believe that others condone their choice to donate are more likely to donate during an epidemic. Gaston and Marc (2013) conducted a behavioral survey of whole blood donors ($N = 252$) intention to donate plasma for the first time. Following a six-month waiting period, 57 donors (22.6%) had given plasma. Godin et al. (2007) followed a similar approach ($N = 2,231$) to build a model of donor decisions to give RBC based on intentions, attitudes, and social norms, to name just a few. Misje et al. (2005) performed a statistical analysis of donor motivations for a blood bank in Norway. With the survey instruments in place, data were collected from blood bank participants over roughly two weeks. In this study, data were acquired once in 2000 ($N_{2000} = 1,029$) and once in 2003 ($N_{2003} = 1,085$). The results showed that benevolent individuals are more likely to give, as compared to those seeking altruistic values. van Dongen (2015) studied the implications of donor motivations on retention. They presented simple solutions to mitigate common negative experiences that influence donor decision to continue giving blood. Charbonneau et al. (2016) surveyed the reasons for donor lapse or reduction in donation frequency ($N = 1,879$) over five years. The results indicated that time constraints (e.g., work obligations) strain donor ability to maintain a frequent donation schedule. Leipnitz et al. (2018) investigated whether offering a health check as an incentive affected donor retention ($N_1 = 53,257, N_2 = 31,522$). The study confirmed that the incentive does improve donor retention.

2.3 Donor Classification

Mostafa (2009) measured blood donor temperament in Egypt using a questionnaire with instruments to record demographics, altruistic values, perceived risks, knowledge, and attitudes from 430 participants. The Support Vector Machine (SVM), Classiffication & Regression Trees (CART), Linear Discriminant Analysis (LDA), and Artificial Neural Network (ANN) algorithms were deployed to classify donors in their dataset as either `donor` or `non-donor`. The best model, ANN, achieves 100% accuracy on the testing dataset. Boonyanusith and Jittamai (2012) conducted a behavioral survey to build a data-frame of donor feelings and attitudes towards donating blood based on responses from 400 participants. In this survey-based study, ANN and the Decision Tree techniques were employed to map members in the population into two discrete groups (i.e., `donor`, `non-donor`). The empirical study showed that the ANN performs reasonably well, achieving $\approx$ 76.25% global accuracy on the dataset. Yeh et al. (2009) compiled a dataset of Taiwanese blood donors using a Recency, Frequency, and Monetary Value (RFM) model as a case study for their proposed RFM solver. This dataset is now publicly hosted by University of California, Irvine (UCI) as the *UCI Blood Transfusion Service Center Data Set*. This dataset contains four features and a target value determining whether the donor gave blood in a particular month. Darwiche et al. (2010) made use of SVM and ANN to predict blood donors in the UCI dataset. The SVM model performed best, achieving 65% sensitivity and 78% specificity on the donor prediction task. Santhanam and Sundaram (2010) applied CART to build decision tree models using the UCI dataset. Khalid et al. (2013) survey common classification problems and proposed solutions in the blood bank sector. They draw attention to the plentiful and successful application of ANN, classification trees, Naíve Bayes, logistic regression, and k-Nearest Neighbors (kNN) for solving classification problems related to blood bank management. Ramachandran et al. (2011) apply the J48 decision tree algorithm – i.e., in the Waikato Environment for Knowledge Analysis (WEKA) software toolkit –to predict recurring donors based on weight, age, sex, and blood group of the donor for a proprietary dataset from the Indian Red Cross Society (IRCS) Blood Bank Hospital. Testik et al. (2012) apply a hybrid methodology of clustering and classification to determine hourly and daily donor arrival patterns using a dataset from Hacettepe University Hospitals' Blood Center. The dataset contains time series data about blood donations to the center. Their model allows blood centers to more intelligently plan staff schedules to match donor trends. Baş et al. (2018) design a mixed integer linear programming to solve a blood donation problem. Namely, their method assigns time slots to blood types using an online system actuated by donor calls to make appointments based on historical data from an AVIS blood center in Milan, Italy. Riley et al. (2007) determined that the existing feature set for modeling donor exclusion – a single feature, donor age – overestimates the available donor pool. They extracted epidemiological features for a set of donors to prove a disparity between the estimated donor pool (i.e., using age) and the actual donor pool (i.e., considering epidemiological exclusions such as HIV).

The vast majority of studies that apply machine learning to blood donor retention use a standardized dataset. Namely, the UCI Blood Transfusion Service Center Data Set, initially presented by Yeh et al. (2009), is frequently studied. The dataset contains four integer-valued features: (1) the number of months since the last donation, (2) the total number of donations, (3) the total amount of blood donated,

and (4) the number of months since the last donation. The final variable, frequently understood as a binary class label, is a Boolean that represents whether the donor gave blood in March 2007. We are not aware of studies utilizing other operation data from blood centers; however, many studies design questionnaires to collect data (Mostafa, 2009; Boonyanusith & Jittamai, 2012).

3 Machine Learning Model of Donor Retention

In this section, we describe a general pipeline for developing machine learning models of donor retention. The pipeline processes relational data provided by BC to predict which donors will give blood during a window of time in the future (e.g., the next month). In doing so, the model can provide BC a means of estimating supply at the donor level. This allows the blood center to make data-driven decisions when designing donor outreach programs by soliciting donors that are likely to be future donors, but have not scheduled new appointments.

Table 5 lists the notations used throughout this section. We separate functional components of our method into a pipeline with four phases depicted in Fig. 1: pre-processing, model development, tuning, and testing. 1) In the *pre-processing* phase, we process data from a relational database into a single table of feature vectors representing blood donors. Then, we define a vector of binary labels determining whether each donor gave blood in a specific time frame. We use sampling techniques to split the data into validation and test sets. Lastly, we balance the distribution of class labels in the validation data to reduce bias towards majority labels. 2) In the *model development* phase, we select and explore different machine learning algorithms on the dataset. 3) In the *tuning* phase, we apply stratified k-fold cross-validation to tune parameters of the computational model (i.e., *hyperparameters*). The configuration of the hyperparameters is known to adversely affect the learning capabilities of the given algorithm, particularly in the case of deep learning. We use deterministic and heuristic search algorithms to find optimal hyperparameter sets in a grid of combinations, namely, grid search and evolutionary computation. 4) Finally, in the *testing* phase, we observe the performance of our models using metrics of accuracy, sensitivity, specificity, and Matthew's correlation coefficient (MCC).

3.1 Pre-processing

In the pre-processing phase, we extract a table of feature vectors representing blood donors from a relational database belonging to BC. We separate this phase into three steps. 1) In *feature engineering*, we aggregate data from tables in the BC database into a single table of features for each donor. 2) Next, we apply stochastic *sampling* to split the table into

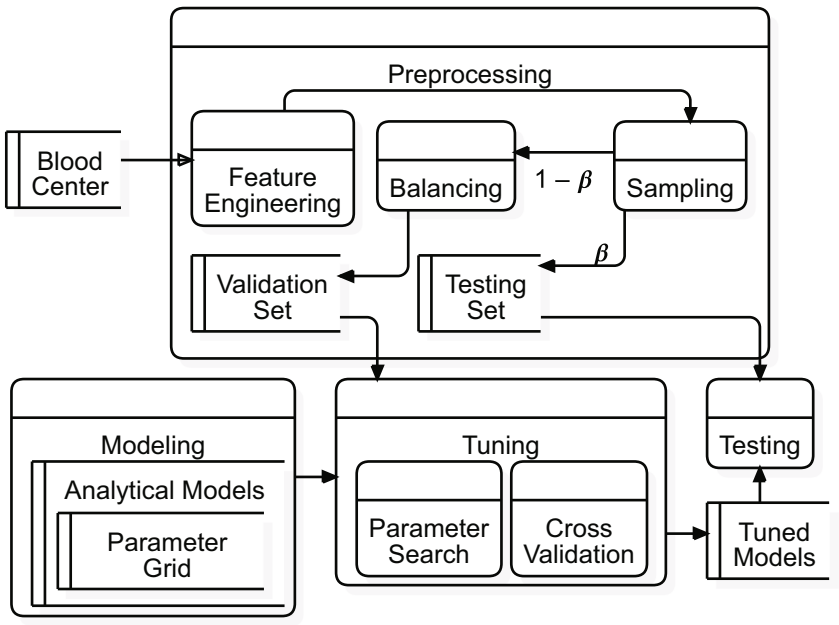

Fig. 1 The pipeline of the phase-based model development. β and $1-\beta$ denote the percentage of data reserved for the testing set and validation set, respectively. Our study applies $\beta = 0.3$

separate validation and testing sets. 3) We apply *balancing* techniques to reduce bias toward majority class labels in the validation data. This balancing step is necessary because the target labels in the data are positive.

During **Feature Engineering**, we start with a table of donor records containing primary demographic data. Using relations in the database, we assess all adjacent tables to pinpoint new features to aggregate into the primary table. We use tables describing: appointment scheduling, donation accreditation groups, donor eligibility, and donor health metrics. After aggregating data into a single dataframe, we convert categorical variables to raw numerical values using either a label encoding or a one-hot matrix. Lastly, we assess any missing values; we discard all records missing mandatory fields and fill non-mandatory missing values with a sentinel value indicating nullity. The BC database contains records ranging from Jan. 1, 2015, to Dec. 20, 2017. We use data from Jan. 1, 2015, to Nov. 1, 2017, to define the feature-based dataset. Then, we extract a set of fulfilled donation records from Nov. 2, 2017, to Dec. 1, 2017, to determine which donors gave blood during November 2017. Donors with at least one complete donation receive a positive label of 1; all others assume a negative label of 0.

To validate and test our models, we divide our single dataset into two sets, namely, *validation* and *testing* sets in the **Sampling** phase. We randomly sample $1 - \beta = 70\%$ of the dataset as the validation data and reserve the remaining $\beta = 30\%$ for the testing purpose.

For any given time window, we observe that few donors in the dataset give blood relative to the total number of registered donors. This results in a strong bias towards the non-donor class when labeling the data for a given time window. Imbalanced data is difficult to model empirically and requires balancing to prevent bias in the model (Johnson & Khoshgoftaar, 2020). We employ a **Balancing** phase to normalize the distribution of class labels. In particular we apply the Synthetic Minority Oversampling Technique (SMOTE), which has been shown to be an effective balancing algorithm (Smiti & Soui, 2020). It is worth noting that SMOTE generates new, synthetic samples belonging to the minority class label to enforce a uniform distribution of class labels in the training data (Chawla et al., 2002). We apply SMOTE to training and validation data, but not testing data to ensure the models test against real-world examples and distributions.

3.2 Model Development

The processed donor retention dataset describes m donors, each bearing n unique features, and a binary label y. We represent each donor as a vector $\boldsymbol{x} \in \mathbb{R}^n$ and their label as an integer $y \in \mathbb{Z}_2$. We commonly refer to the dataset as the matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ and the ground truth labels as the vector $\boldsymbol{y} \in \mathbb{Z}_2^m$. Solutions to the donor retention problem involve the approximation of a function $f^* : \mathbb{R}^n \rightarrow \mathbb{Z}_2$ mapping donors $\boldsymbol{x}$ to labels y (Goodfellow et al., 2016). We approximate f^* as $\hat{y} = f\{\boldsymbol{x}\}$ where $\hat{y}$ is the predicted binary label for a donor $\boldsymbol{x}$ from some approximate mapping $f\{\boldsymbol{x}\}$. In this study, we apply a variety of machine learning algorithms shown to work well in the domain of health-care: kNN, CART, and ANN (Tomar & Agarwal, 2013). We also apply distinct ensembles of CART: Random Forest (RF), and Gradient Boosting (GB). Table 1 describes the models in terms of their respective properties.

3.3 Tuning

We identify a grid of valid hyperparameter combinations for each model and use search algorithms to find the optimal set. In the case of kNN, we apply a distributed brute force search over the 300 combinations of parameters. However, the remaining models embody hyperparameter spaces with at least millions of combinations. To efficiently probe the hyperparameter search space, a hill-climbing algorithm based on heuristics of evolution is applied. Namely, we utilize an Evolutionary Algorithm to approximate the optimal hyperparameter set in 100 candidate evaluations. Because the neural net requires extensive and in-depth tuning, we use ad-hoc experimentation to tune this particular model. Irrespective of the search technique, we use stratified k-fold cross-validation on the dedicated validation dataset to generate a scalar metric for comparison. We hold $k = 3$ constant throughout all of our experiments.

Table 1 A comparison of machine learning models applied in this study

Model	Ensemble	Instance–based	Gradient–based	Feature Selector	Parameters
kNN	×	✓	×	×	3
CART	×	×	×	✓	7
RF	✓	×	×	✓	9
GB	✓	×	✓	✓	9
ANN	✓	×	✓	✓	∞

Because the engineered dataset undergoes no dimensionality reduction or feature selection before model development, we expect decision tree–based classifiers to better adapt to the problem space. Furthermore, we anticipate the GB and RF ensembles to produce the most reliable predictions by marginally boosting the performance of single decision trees. Decision trees inherently select high priority features thanks to a greedy choice over entropy reduction at each node. On the contrary, *k*NN and ANN use the global feature space resulting in excess noise in the predictions. Weak features can have a negative impact on *k*NNs and ANNs ability to form stable decision boundaries (Domeniconi C & Gunopulos D, 2001; Tirelli & Pessani, 2011). It is worth noting that ANNs do weight features automatically at the input layer. However, noisy features can poison the gradient before an optimizer converges on weights that remove this input noise.

3.4 Testing

We introduce notations $d_{p\to p'}$, $d_{p\to n'}$, $d_{n\to p'}$, and $d_{n\to n'}$ to define performance metrics in the testing phase. Let $d_{p\to p'}$ be the number of donors classified as giving blood in the future that do donate; $d_{p\to n'}$ is the number of donors classified as donating blood that do not give; $d_{n\to p'}$ is the number of donors classified as not giving blood, but do; $d_{n\to n'}$ is the number of donors classified as not giving blood that indeed do not. We define the total number of positive and negative predictions as P and N, respectively. Likewise, P' and N' represent the total number of positive and negative truths. Fig. 2, a confusion matrix, summarizes the notation. Detailed symbols and annotations are provided in Appendix A.

We use notation from the confusion matrix in Fig. 2 to define scalar performance metrics of accuracy, sensitivity, specificity, and MCC. Equation 1 describes A, the *accuracy*, as the percentage of donors that are correctly identified as giving or non-giving among all the donors.

$$A = \frac{d_{p\to p'} + d_{n\to n'}}{P + N} \tag{1}$$

T represents *sensitivity*, or true positive rate, which is the fraction of correctly identified giving donors among all the donors that do give. As such, Eq. 2 expresses T.

$$T = \frac{d_{p\to p'}}{P'} \tag{2}$$

F denotes the *specificity*, or true negative rate, which is the fraction of correctly identified non-giving donors among all the donors that do not give. Thus, we can describe F using Eq. 3.

$$F = \frac{d_{n\to n'}}{N'} \tag{3}$$

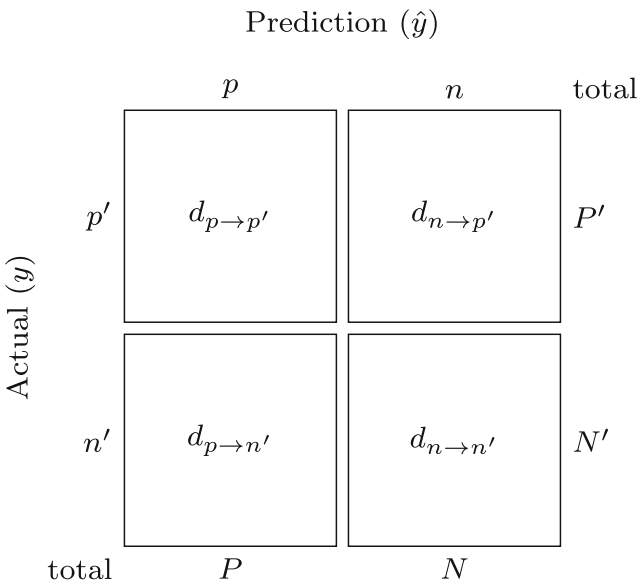

Fig. 2 The confusion matrix for evaluating models. $d_{p\to p'}$ is the number of donors classified as giving blood in the future that do in fact donate; $d_{p\to n'}$ is the number of donors classified as donating blood that do not actually give; $d_{n\to p'}$ is the number of donors classified as not giving blood, but actually do; $d_{n\to n'}$ is the number of donors classified as not giving blood that indeed do not

M is the *MCC*, which is the correlation coefficient between the predicted values and the actual values (Matthews, 1975). We define M in Eq. 4. The MCC produces an output in the range $[-1, 1]$ where a value of 1 indicates perfect predictions, 0 implies no better than random guess, and -1 means entirely inverted predictions. It is worth noting that the MCC is shown to produce a balanced measure despite imbalanced class labels (Boughorbel et al., 2017).

$$M = \frac{d_{p\to p'} \times d_{n\to n'} - d_{p\to n'} \times d_{n\to p'}}{\sqrt{P \times P' \times N \times N'}} \tag{4}$$

4 Results

4.1 Machine Learning Models and Validation

To model donor retention, we first aggregate data from various tables in a relational database to a single table where each sample represents a donor. We use timestamps in the database to select records from Jan. 1, 2015, through Nov. 1, 2017. Where necessary, we apply a map-reduce technique to resolve one-to-many relationships within the database. We map categorical variables to integers using a simple one-to-one encoding. The final dataset – described by Table 6 (Appendx B) with correlation matrix visualized in Fig. 4 and distributions shown in Fig. 5 (Appendx B) – contains $n = 51$ unique features characterizing each of the

Table 2 Training results of the best hyperparameter configuration for each model sorted in descending row-order by MCC

Model	Accuracy (A)	Sensitivity (T)	Specificity (F)	MCC (M)
RF	0.994	0.993	0.995	0.988
GB	0.994	0.993	0.996	0.988
DT	0.986	0.985	0.988	0.973
kNN	0.973	0.992	0.955	0.947
ANN	0.937	0.921	0.953	0.874
Random	0.5	0.5	0.501	0.0

$m \approx 6.3 \cdot 10^4$ donors. For each donor, we define a binary label that represents whether the donor gave blood during November of 2017; only $2.3e3$ (3.6%) people donated. We select the last month as the target period because the engineered dataset derives from a snapshot of an operational database that contains timestamp data. We split the dataset into validation and testing sets using a testing set size of $\beta = 0.3$. Then, we employ SMOTE to balance the validation data producing a uniform distribution of class labels for training. It is worth noting, correlated features are not removed from the dataset before the model development phase. We found that removing the correlated features hurt the performance of the machine learning algorithms.

With the data mentioned above, we present a selection of machine learning models, each bearing a unique technique for fitting data. We collect metrics for evaluating the performance of individual models: accuracy, sensitivity, specificity, and MCC. In this case, accuracy captures the rate of correct label identification among all the labels, sensitivity refers to the prediction rate of donors that will donate, and specificity measures the rate of correct identification of non-donors. Because the class labels are imbalanced in the testing data, we measure MCC to provide an unbiased assessment of models across both labels.

We utilize a variety of techniques to tune the hyperparameters of each model. In the case of kNN, we use a brute force grid search over 300 parameter combinations to find the optimal set. CART, GB, and RF classifiers present millions of potential combinations, yielding brute force search infeasible. Instead, we apply an Evolutionary Algorithm to find nearly optimal hyperparameters sets in $\approx 1,500$ candidate solutions. We find ad-hoc experimentation the most effective method for tuning the neural network due to model complexity and hyperparameter instability.

Table 2 lists the training metrics for each algorithm tested. The algorithms are listed in descending order by MCC such that more competent models are at the top. All of the studied algorithms are capable of fitting the training dataset in terms of the studied metrics. In particular, the performance difference between GB and RF classifiers is practically immeasurable. As ensembles of multiple decision trees, it is intuitive that GB and RF classifiers would marginally improve the performance of a basic decision tree classifier, which can be observed across all tested metrics. Both the k-nearest neighbors and artificial neural net classifiers fit the training data to a reasonable extent, but fail to achieve comparable performance to the decision tree–based classifiers.

Table 3 exhibits the training accuracy of each model for each of the 3 folds in the cross-validation scheme. The accuracy remains stable among each fold for a given model, indicated by the low standard deviation in accuracy. This suggests that the underlying dataset properly conforms to the i.i.d. assumption, i.e., is independent and identically distributed.

Using naive grid search, we were able to determine that the most effective parameterization of k-nearest neighbors on the data involved using a single neighbor (i.e., $k = 1$) and the Canberra distance metric with no instance weighting. An evolutionary algorithm was used to optimize the decision

Table 3 The training accuracy of each of the 3 folds in the training phase

Model	Fold 1	Fold 2	Fold 3	Std Dev (σ)
GB	0.994	0.994	0.994	0.0
RF	0.994	0.994	0.993	0.0006
DT	0.987	0.987	0.987	0.0
kNN	0.975	0.973	0.973	0.0012
ANN	0.854	0.939	0.947	0.0515
Random	0.502	0.502	0.497	0.0029

tree–based classifiers due to prohibitively large search space. In the case of the decision tree, it was determined that the entropy criterion was more effective than the Gini index. Additionally, limiting the depth of the tree to $b = 30$ was found to be effective, though no other limiting mechanism proved effective. For the random forest ensemble, the entropy mechanism also showed itself to be more productive than the Gini index. A depth limit of $b = 40$ was found to be effective, along with a limit on the number of features for each estimator in the ensemble, namely, the square root of the number of total features. It was also useful to limit the minimum number of samples for a split in the tree to 9. Although more computationally demanding, it was found that more estimator improve the performance of the random forest classifier. In particular, 11 estimators are used by the most optimal model, which is the upper bound in our search. The gradient boosting trees algorithms optimal parameter set was found to include 9 estimators and a learning rate of 0.1. Very similar to the random forest classifier, a depth limit of $b = 30$, a feature cap of the square root of the total number of features, and a minimal tree split sample quantity of 8 produced an effective model.

For the neural network model, we determined by ad-hoc experimentation – opposed to grid search or an evolutionary algorithm – that an architecture using processing blocks that consisted of (1) batch normalization, (2) a hyperbolic tangent activation, (3) a densely connected layer, and (4) a dropout layer with a drop rate of 0.2 was effective on the training data. The final model consisted of three of these processing blocks with 64, 32, and 16 dense units respectively. A final dense layer maps the activation map to a single logit that passes through a sigmoidal function to produce a binary class label (assuming a decision threshold of 0.5).

4.2 Testing Results

Table 4 summarizes the results of the testing phase. We observe that each model achieves an accuracy of at least 93%, suggesting that the models fit the problem space well. However, small variations of accuracy seem to correlate to significant variations of sensitivity. For instance, GB is only 2.7% more accurate than ANN, but is 19.1% more sensitive. This result suggests that accuracy is a misleading metric. We notice a similar trend when assessing specificity; each model scores at least 96.0%, but the less specific models are far less sensitive.

These observations likely result from the imbalanced distribution of class labels (i.e., "donor," "non-donor") in the validation and testing data. Because the number of non-donors far exceeds the number of donors, models that output the mode label can quickly achieve high accuracy and perfect specificity. Conversely, models that output the inverse label would earn poor accuracy, but a perfect sensitivity. As such, these three metrics obfuscate the performance of these models and provide no sole means of comparison.

We note that MCC produces a fair and unitary performance measure of each model in the presence of imbalanced data (Boughorbel et al., 2017). As such, we sort the results in Table 4 in descending order by MCC. In this way, we confirm that decision tree–based methods – more specifically CART, RF, and GB – far outperform both ANN and *k*NN. Ultimately, RF stands out as the most reliable model achieving a confident 99.7% specificity, while maintaining a high 80.1% sensitivity.

4.3 Feature Importances

A constructed decision tree inherently prioritizes features at each node based on a greedy reduction in split criterion. As such, CART and its ensembles provide a natural mechanism to assess which features determine whether donors will give blood. For any given feature, we can compute the importance as the normalized total reduction in entropy brought on by splits involving the feature. Figure 3a and b depict importance measures based on the best RF and GB classifier, respectively. Although both models produce similar evaluation metrics in Table 4, their underlying decision process differs. RF places a substantial importance on two features of `nextDonationDate`, `nextContactDate`. Logically, these fields provide an explicit mapping to future donations. GB produces a more diverse distribution of features in its importance map. It

Table 4 Testing results of the best hyperparameter configuration for each model sorted in descending row-order by MCC

Model	Accuracy (A)	Sensitivity (T)	Specificity (F)	MCC (M)
RF	0.99	0.801	0.997	0.851
GB	0.989	0.79	0.997	0.844
DT	0.983	0.807	0.99	0.776
ANN	0.962	0.599	0.976	0.524
kNN	0.937	0.168	0.967	0.133
Random	0.504	0.477	0.505	-0.007

heavily favors the same three features as RF, but reserves more importance for a small subset of demographic and scheduling information. Both importance maps suggest that simple logistical and demographic data determine retention more than health diagnostics. SHAP values were computed for Decision Tree, Gradient Boosting, and Random Forest classifiers. However, we found the results to be qualitatively identical to the existing result using importance measures. As such, SHAP values are not presented in Table 4.

5 Discussion

In Section 4, we presented the results of our machine learning pipeline on the BC data. Predicting donor retention can provide blood centers a means of reducing waste in outreach programs by targeting the donors that are more likely to return to give blood in the following month. By pruning donors that are unlikely to return from outreach lists, unnecessary cost due to excessive e-mail, phone, and mail promotions can be reduced. In a time where blood supply and transfusion are both decreasing, such cost reduction mechanisms are highly prudent for blood centers. Although the focus of these models is on donors that exist in a blood center dataset, – and have thus donated at least once, – the application extends to the case of donor recruitment as well. If a model of returning donors for a particular month predicts a lower than average number of returning donors, this may be an indicator to the blood center to engage in donor recruitment programs to bring new donors into the system.

This study aimed to answer three research questions related to machine learning methodology for blood supply chain management. Following a data mining phase where a tabular classification dataset was constructed from an operational database belonging to BC, a set of machine learning models was constructed and tested to determine the algorithm that produced the highest testing results in terms of four metrics. The best approach in this study, RF, achieved a sensitivity of 80% while maintaining a high specificity of 99%. Although RF is quantitatively the best, we argue that other CART ensembles, like RF and GB, compete with this score. CART fits the space well for two reasons, namely, (1) a small subset of robust features and (2) an inherent ability to filter out noise. Importance plots from component trees of GB and RF ensembles confirm that three features agreeably comprised the majority of the split importance measures for these models. The resulting increase in performance seen in GB and RF models compared to CART seemingly emerged from the adaptation to more complex and nuanced decision boundaries among non-modal donors. Although the importance readings revealed that the majority of features provided little to no insight into whether a donor will give blood, these noise features have no impact on CART or associated ensembles. Because CART is greedy, it selects the feature with the highest importance at each split node. As such, CART–based methods perform well in areas where the feature space may contain excess noise. This is not the case for the *k*NN model, which relies on the unweighted distance between vectors according to a metric like Euclidean distance. This distance metric can be easily influenced by excess noise that produces too much variance in the data. Additionally, ANN can struggle to adapt to cases where there is excess feature noise due to gradient poisoning and related problems. Future work may address whether subsets of the feature space produce higher quality results from *k*NN and ANN models. Although prior studies

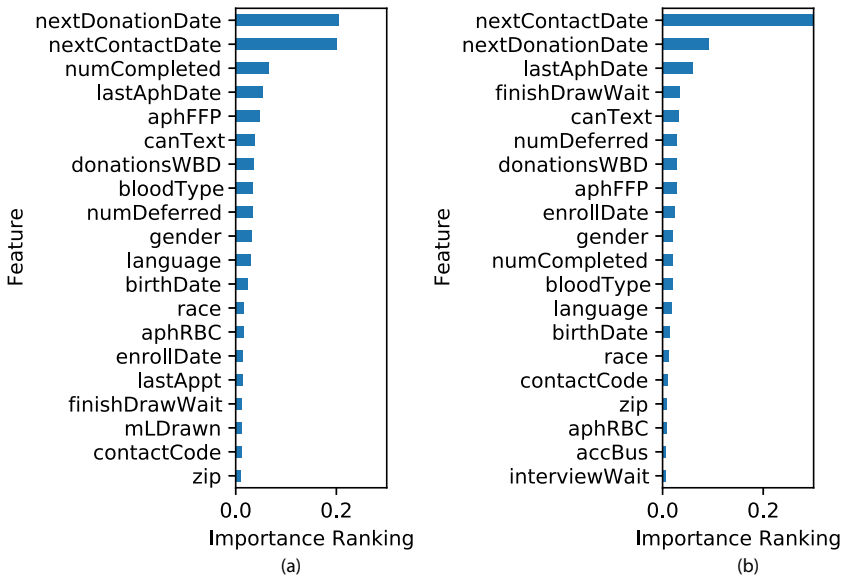

Fig. 3 Importance rankings of dataset features based on trained RF (3a) and GB (3b) classifiers. Both models place high importance on the `nextContactDate`, `nextDonationDate`, and `lastAphDate` features. Although the models agree on the three most important features, GB spreads the importance across more variables

have seen success from SVM models, the SVM performed no better than random guess when applied to our dataset. As such, data surrounding the SVM has been omitted from the paper. It is worth noting, the implementation of GB applied in this study is the XGBoost implementation (Chen & Guestrin, 2016).

Most machine learning algorithms are parametric and have complex hyperparameter spaces that can be challenging to tune. In this study we addressed the tuning of hyperparameters using three distinct methods depending on the model. In the case of kNN, we employed a deterministic grid search using a Spark cluster to distribute the workload. We were able to determine that using $k = 1$, i.e., one neighbor, was the most effective. Because there may be noise in the input features, this result is logical because noise will be magnified as k increases. The tree-based methods, – CART, GB, and RF, – have prohibitively large search spaces, rendering grid search plainly infeasible without sacrificing a significant amount of sample resolution. To overcome this limitation, we applied an evolutionary algorithm to navigate the search space stochastically based on well-founded heuristics of evolution. This allows us to calculate hyperparameter sets that produce locally optimal results without incurring the prohibitive overhead of brute force hyperparameter optimization. We discovered that limiting the breadth, depth, split quantity, etc. of the trees that are built produced higher quality models than unconstrained trees. With regards to testing metrics, this is a relatively intuitive conclusion as such restrictions have a regularizing effect that prevents the models from over-fitting to the training data. In the case of the ANN, the model was tuned in terms of the number of layers and the number of neurons per layer. We found that three layers with 64, 32, and 16 units each to produce the best testing metrics. Additional network depth or breadth caused over-fitting to the training data and did not encourage a better result in terms of testing metrics.

Tree-based classifiers have a structure that naturally forms a quantitative hierarchy of concepts over the feature space. This structure can be utilized to analyze the importance of different features to the underlying predictions. Assuming a model that produces high quality predictions, such analysis can reveal the features that are most strong predictors. We studied the feature importances of trained RF and GB models to gain an understanding of the features in the data that are have a strong influence on donor prediction. The two strongest features that determined whether a donor would give blood in the following month were the `nextDonationDate` and `nextContactDate` features. The `nextDonationDate` represents when a particular donor is actually scheduled to give blood in the future, which is an intuitively good indicator of retention. The `nextContactDate` variable describes when the blood center will next contact the donor to solicit a donation. The strength of this variable suggests that donors subscribed to receive solicitations through the system actively respond to them through routine donations. This suspicion is further evidenced by the strength of the `canText` feature. Aggregate data such as the total amounts of donated blood and number of deferrals were features of slight importance that varied somewhat between the GB and RF models. Demographic information such as gender, race, zip-code, etc. had a low importance when predicting donors. Future work may consider studying feature ablation to determine nuanced relationships between subsets of weaker features. I.e., removing `nextDonationDate` and `nextContactDate` from the dataset will encourage the tree-based classifiers to find patterns in the weakened signal.

Yeh et al. presented the first dataset of blood donors, describing some $\approx$ 700 randomly sampled patrons of a Taiwanese transfusion center (Yeh et al., 2009). This dataset, now hosted by UCI Machine Learning repository (Dheeru & Karra Taniskidou, 2017), receives great attention from academic communities as a research platform for machine learning methods. The relative size of this dataset and the different metrics used to quantify performance illicit a poor comparison to our study. To the best of our knowledge, the dataset of donors presented in this study is the largest to date, comprising some 51 features for each of the $6.3e4$ donors. Furthermore, we found no study assessing the importance measures of features in any donor database.

As a final note, we mine the tabular donor retention dataset from a historical archive spanning two years and ten months for BC. We predict donor retention for a time window of one month, i.e., the model answers the question: "which donors will return and give blood the following month?" The usage of these time spans is consistent with how blood centers practically operate at the business level. Assessing the efficacy of using these particular windows is beyond the scope of this study. However, future studies can assess the efficacy of using sliding windows of differing lengths for predicting donor retention. Future studies could be designed with larger longitudinal data so multiple planning horizons could be included.

6 Conclusion

Bearing comprehensive operational data from a regional blood center, we studied machine learning models for predicting blood donors. We processed the large donor dataset and defined a one-month window to predict whether or not those donors gave blood. In doing so, we

discovered that ensembles of decision trees – GB and RF – performed well on this problem, dwarfing the kNN, ANN, and SVM approaches. We showed that the decision trees favor a consistent set of donor features related to scheduling, planning, and aggregate donor metrics. The models developed in this work can be utilized by blood centers to build better outreach programs through data-driven decision-making. Blood centers can apply themodels to identify donors that may be likely to donate, but have not schedule an appointment and solicit them for donations. Alternatively blood centers can save costs by identifying unlikely donors to prune from outreach programs, thus saving unnecessary mail and phone calls.

Appendix A: Symbols and Annotations

Table 5 Symbols and annotations

Symbol	Annotation
β	The percentage of data reserved for the testing set
m	The number of donors in the final dataset
n	The number of unique features per donor
$\boldsymbol{x}$	An n-dimensional vector representing a donor
y	A label for a donor determining if they donated in the future
$\boldsymbol{X}$	An $m \times n$ dimensional matrix of donor feature vectors
$\boldsymbol{y}$	An m-dimensional vector of donor labels (y)
f^*	An optimal mapping from donors ($\boldsymbol{x}$) to labels (y)
f	An approximate mapping from donors ($\boldsymbol{x}$) to predicted labels ($\hat{y}$)
$\hat{y}$	A predicted label from some f based on some $\boldsymbol{x}$
$d_{p\rightarrow p'}$	The number of correctly identified future donors
$d_{p\rightarrow n'}$	The number of incorrectly identified future donors
$d_{n\rightarrow p'}$	The number of incorrectly identified future non-donors
$d_{n\rightarrow n'}$	The number of correctly identified future non-donors
A	The accuracy of a model's predictions
T	The sensitivity of a model's predictions (the true positive rate)
F	The specificity of a model's predictions (the true negative rate)
M	Matthew's correlation coefficient (MCC) of a confusion matrix

Appendix B: Dataset Features

Table 6 The engineered features of the donor retention dataset

Feature	Description
accBus	The number of unique businesses accredited by the donor.
accEdu	The number of unique colleges accredited by the donor.
accGov	The number of unique government entities accredited by the donor.
accOrg	The number of unique organizations accredited by the donor.
accRel	The number of unique churches accredited by the donor.
aphFFP	The amount of fully frozen plasma given by the donor in their last apheresis procedure.
aphPLT	The amount of platelets given by the donor in their last apheresis procedure.
aphRBC	The amount of red blood cells given by the donor in their last apheresis procedure.
apptsAphPTA	The number of apheresis appointments where the donor gave platelets.
apptsSAP	The number of appointments where the donor gave source plasma.
apptsWBD	The number of appointments where the donor gave whole blood.

Table 6 (continued)

Feature	Description
birthDate	The birthdate of the donor.
bloodType	The donor's tested blood type.
canText	A flag determining whether to send text messages to the donor.
city	The city of the donor's primary residence.
contactCode	A code determining how/whether to contact the donor.
didRinseback	A flag determining whether rinseback was used in the last apheresis procedure for the donor.
donationsAphDPA	The number of apheresis donations where the donor gave plasma.
donationsAphDRA	The number of apheresis donations where the donor gave red blood cells.
donationsAphPPA	The number of apheresis donations where the donor gave platelets / plasma.
donationsAphPPR	The number of apheresis donations where the donor gave platelets / plasma / red blood cells.
donationsAphPRA	The number of apheresis donations where the donor gave platelet / red blood cells.
donationsAphPTA	The number of apheresis donations where the donor gave platelets.
donationsSAP	The number of donations where the donor gave source plasma.
donationsSUR	The number of donations where the donor gave single alyx.
donationsWBD	The number of donations where the donor gave whole blood.
enrollDate	The original enrollment date of the donor at a blood center.
finishDrawWait	The average amount of time spent by the donor having blood actively drawn.
gallonsGiven	The number of gallons given by the donor.
gender	The gender of the donor.
interviewWait	The average amount of time spent by the donor between check-in and an interview.
isDonating	A flag denoting whether the donor wants to donate.
language	The primary langauge spoken by the donor.
lastAphDate	The date of the donors last apheresis appointment.
lastAppt	The date of the donor's last appointments at a blood center.
lastContactDate	The date the donor was last contacted by telerecruiting.
lastDonationDate	The date the donor last gave blood.
mLDrawn	The total number of milliliters of red blood cells donated by the donor.
nextContactDate	The date when the donor should be contacted next.
nextDonationDate	The date when the donor is eligible to donate.
numCompleted	The total number of donations completed by the donor.
numDeferred	The total number of donations by the donor that were deferred.
numWalkedOut	The total number of donations that the donor walked out from.
qualCnt	The qualifying count of the donor before their last apheresis donation.
qualCntBasis	A code denoting the basis of the qualifying count.
race	The race of the donor.
startDrawWait	The average amount of time spent by the donor between an interview and a draw to start.
state	The state of the donor's primary residence.
suffix	The suffix of the donor's name.
totalDonations	The number of previous donations made at other blood centers.
zip	The zip code of the donor's primary residence.

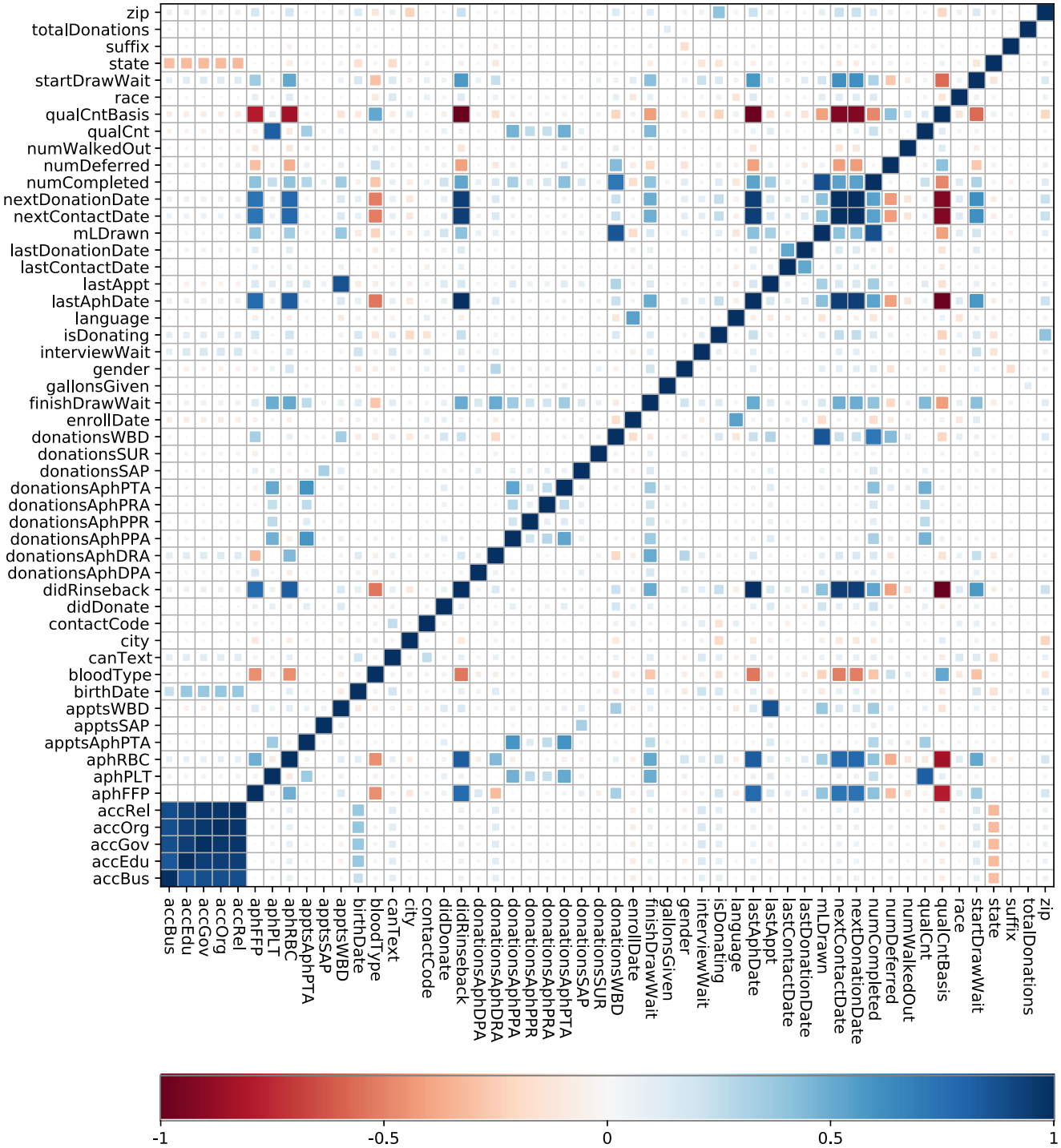

Fig. 4 A heatmap visualization of the correlation matrix of the engineered dataset described by Table 6. The color (luminance) of a cross-section, as well as the area, measures the Pearson correlation between variables in the dataset

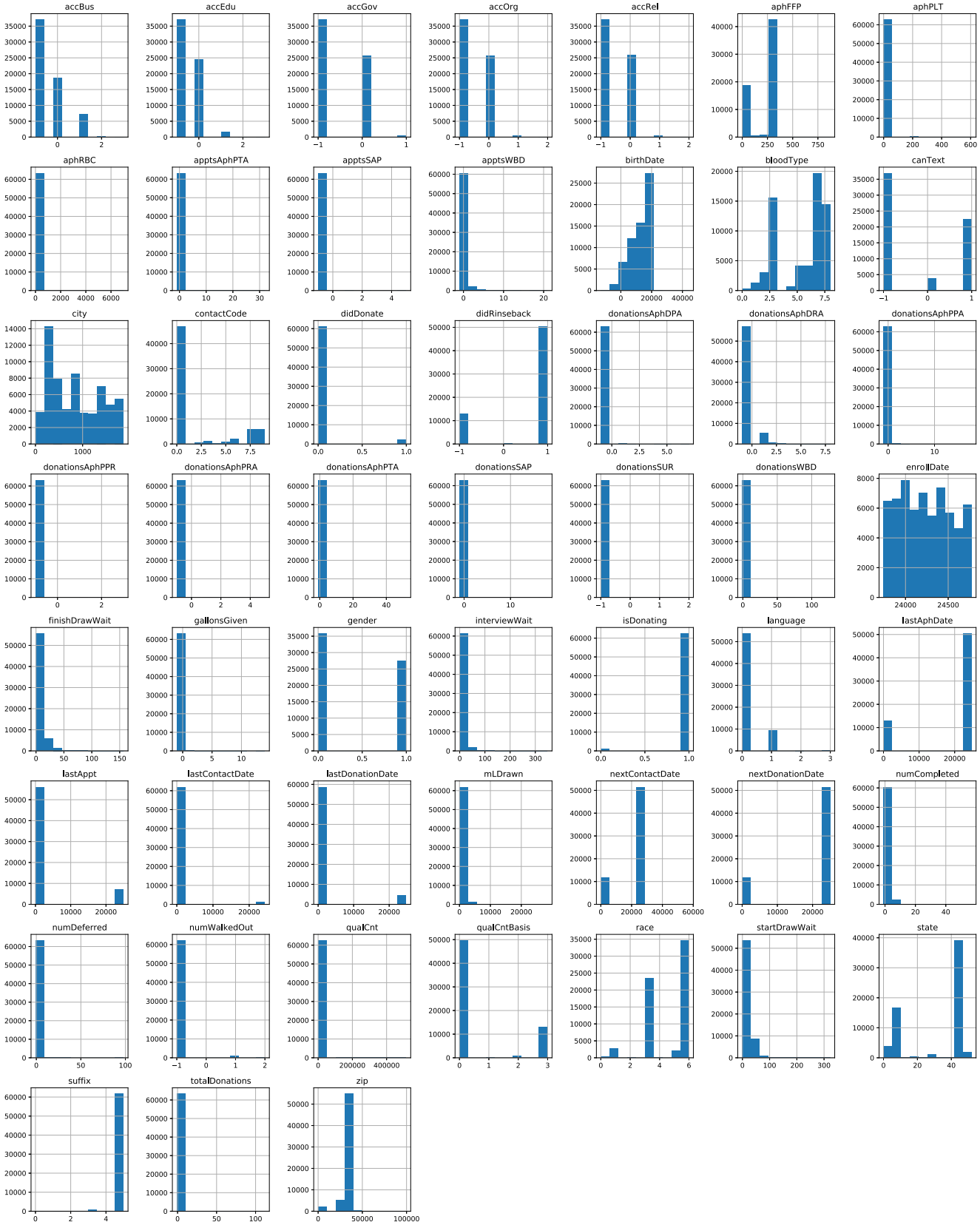

Fig. 5 Histogram plots illustrating the distributions of variables in the dataset

Glossary

BC	Blood Center (Name blinded)
ARC	The American Red Cross
ANN	Artificial Neural Network
CART	Classification & Regression Trees
CBA	Classification Based Association
CDC	The Center for Disease Control
DRG	Diagnosis Related Groups
FEMA	The Federal Emergency Management Agency
FDA	The Food and Drug Administration
GB	Gradient Boosting
IRCS	Indian Red Cross Society
*k*NN	*k*-Nearest Neighbors
LDA	Linear Discriminant Analysis
MCC	Matthew's correlation coefficient
NBCUS	The National Blood Collection and Utilization Survey
PBM	Patient Blood Management
RBC	Red Blood Cell
RF	Random Forest
RFM	Recency, Frequency, and Monetary Value
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
UCI	University of California, Irvine
US	United States

Acknowledgements Xiao Qin's work is supported by the U.S. National Science Foundation under Grants IIS-1618669, OAC-1642133, CCF-0845257.

References

Abbasi, B., & Hosseinifard, S. Z. (2014). On the issuing policies for perishable items such as red blood cells and platelets in blood service. *Decision Sciences*, *45*(5), 995–1020.

Abbasi, B., Vakili, G., & Chesneau, S. (2017). Impacts of reducing the shelf life of red blood cells: a view from down under. *INFORMS Journal on Applied Analytics*, *47*(4), 336–351.

Baş, S., Carello, G., Lanzarone, E., & Yalçındağ, S. (2018). An appointment scheduling framework to balance the production of blood units from donation. *European Journal of Operational Research*, *265*(3), 1124–1143.

Beliën, J., & Forcé, H. (2012). Supply chain management of blood products: a literature review. *European Journal of Operational Research*, *217*(1), 1–16.

Boonyanusith, W., & Jittamai, P. (2012). Blood donor classification using neural network and decision tree techniques. In *Proceedings of the world congress on engineering and computer science*, (Vol. 1 pp. 499–503).

Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PLOS ONE*, *12*(6), 1–17.

Charbonneau, J., Cloutier, M. S., & Carrier, É. (2016). Why do blood donors lapse or reduce their donation's frequency? *Transfusion Medicine Reviews*, *30*(1), 1–5.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Darwiche, M., Feuilloy, M., Bousaleh, G., & Schang, D. (2010). Prediction of blood transfusion donation. In *2010 Fourth International Conference on Research Challenges in Information Science (RCIS)* (pp. 51–56).

Dheeru, D., & Karra Taniskidou, E. (2017). UCI machine learning repository. http://archive.ics.uci.edu/ml.

Domeniconi C, & Gunopulos D (2001). Adaptive nearest neighbor classification using support vector machines. In *Proceedings of the 14th International conference on neural information processing systems: natural and synthetic, MIT Press, Cambridge, MA, USA, NIPS'01* (pp. 665–672).

van Dongen, A. (2015). Easy come, easy go. retention of blood donors. *Transfusion Medicine*, *25*(4), 227–233.

Ellingson, K. D., Sapiano, M. R. P., Haass, K. A., Savinkina, A. A., Baker, M. L., Chung, K. W., Henry, R. A., Berger, J. J., Kuehnert, M. J., & Basavaraju, S.V. (2017). Continued decline in blood collection and transfusion in the United States - 2015. *Transfusion*, *57*(Suppl 2), 1588–1598.

Gaston, G., & Marc, G. (2013). Predicting first lifetime plasma donation among whole blood donors. *Transfusion*, *53*(S5), 157S–161S.

Godin, G., Conner, M., Sheeran, P., Bélanger-Gravel, A., & Germain, M. (2007). Determinants of repeated blood donation among new and experienced blood donors. *Transfusion*, *47*(9), 1607–1615.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge: MIT Press. http://www.deeplearningbook.org.

Gupta, A., Deokar, A., Iyer, L., Sharda, R., & Schrader, D. (2018). Big data & analytics for societal impact: Recent research and trends. *Information Systems Frontiers*, *20*(2), 185–194.

Johnson, J. M., & Khoshgoftaar, T. M. (2020). The effects of data sampling with deep learning and highly imbalanced big data. *Information Systems Frontiers*, *22*(5), 1113–1131.

Kamyabniya, A., Lotfi, M. M., Naderpour, M., & Yih, Y. (2018). Robust platelet logistics planning in disaster relief operations under uncertainty: a coordinated approach. *Information Systems Frontiers*, *20*(4), 759–782.

Khalid, N. S. C., Burhanuddin, M., Ahmad, A., & Ghani, M. (2013). Classification techniques in blood donors sector–a survey. In *E-Proceeding of Software Engineering Postgraduates Workshop (SEPoW)*.

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information systems frontiers*, *19*(2), 267–283.

Leipnitz, S., de Vries, M., Clement, M., & Mazar, N. (2018). Providing health checks as incentives to retain blood donors — evidence from two field experiments. *International Journal of Research in Marketing*, *35*(4), 628–640.

Masser, B. M., White, K. M., Hamilton, K., & McKimmie, B.M. (2011). An examination of the predictors of blood donors' intentions to donate during two phases of an avian influenza outbreak. *Transfusion*, *51*(3), 548–557.

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA,) - Protein Structure*, *405*(2), 442–451.

Miah, M. (2020). Study of blood donation campaign communication methods and attributes of donors: a data analytics approach. *International Journal of Healthcare Management*, *0*(0), 1–11.

Misje, A. H., Bosnes, V., Gåsdal, O., & Heier, H.E. (2005). Motivation, recruitment and retention of voluntary non-remunerated blood donors: a survey-based questionnaire study. *Vox Sanguinis*, *89*(4), 236–244.

Mostafa, M. M. (2009). Profiling blood donors in egypt: a neural network analysis. *Expert Systems with Applications*, *36*(3, Part 1), 5031–5038.

Mulcahy, A., & Health, R. (2016). Toward a sustainable blood supply in the United States: an analysis of the current system and

alternatives for the future. Research report (Rand Corporation), RAND Corporation.
Osorio, A. F., Brailsford, S. C., & Smith, H.K. (2015). A structured review of quantitative models in the blood supply chain: a taxonomic framework for decision-making. *International Journal of Production Research*, *53*(24), 7191–7212.
Pierskalla, W. P. (2005). *Operations research and health care* (Vol. 70, pp. 103–145). Boston: Springer. chap Supply Chain Management of Blood Banks.
Ramachandran, P., Girija, N., & Bhuvaneswari, T. (2011). Classifying blood donors using data mining techniques. *International Journal of Computer Science Engineering & Technology, 1*(1).
Riley, W., Schwei, M., & McCullough, J. (2007). The united states' potential blood donor pool: estimating the prevalence of donor-exclusion factors on the pool of potential donors. *Transfusion*, *47*(7), 1180–1188.
Santhanam, T., & Sundaram, S. (2010). Application of CART algorithm in blood donors classification. *Journal of computer Science*, *6*(5), 548.
Smiti, S., & Soui, M. (2020). Bankruptcy prediction using deep learning approach based on borderline SMOTE. *Information Systems Frontiers*, *22*(5), 1067–1083.
Testik, M. C., Ozkaya, B. Y., Aksu, S., & Ozcebe, O.I. (2012). Discovering blood donor arrival patterns using data mining: a method to investigate service quality at blood centers. *Journal of Medical Systems*, *36*(2), 579–594.
Tirelli, T., & Pessani, D. (2011). Importance of feature selection in decision-tree and artificial-neural-network ecological applications. alburnus alburnus alborella: a practical example. *Ecological Informatics*, *6*(5), 309–315.
Tomar, D., & Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, *5*(5), 241–266.
Whitaker, B. I., Henry, R. A., & Hinkins, S. (2013). AABB Blood collection, utilization, and, patient blood management survey report. Tech. rep., American Assocation of Blood Banks (AABB).
Yeh, I. C., Yang, K. J., & Ting, T.M. (2009). Knowledge discovery on rfm model using bernoulli sequence. *Expert Systems with Applications*, *36*(3, Part 2), 5866–5871.

Mr. Christian Kauten received the B.S and M.S degrees in Software Engineering from Auburn University in 2017 and 2019, respectively. He is currently a Ph.D. candidate and graduate research assistant at Auburn University studying Software Engineering. His current research interests include machine learning, digital signal processing, and computer vision.

Professor Ashish Gupta is an Associate Professor of Analytics in Raymond J. Harbert College of Business at the Auburn University. He has a Ph.D. in MSIS from Spears School of Business at Oklahoma State University. His research interests are in the areas of data analytics, healthcare informatics, sports analytics, organizational and individual performance. His recent articles have appeared in journals such as European Journal of Information Systems, Decision Sciences Journal, European Journal of Operations Research, MIT Sloan Management Review, Risk Analysis, Journal of American Medical Informatics Association, Journal of Biomedical Informatics, IEEE Transactions, Information Systems Journal, Decision Support Systems, Information Systems Frontiers, Communications of the Association for Information Systems, etc. He has published 5 edited books. His research has been supported by various grant agencies such as THEC, DHS, and several private organizations.

Professor Xiao Qin received the B.S. and M.S. degrees in Computer Science from the Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree in Computer Science from the University of Nebraska-Lincoln, Lincoln, in 1992, 1999, and 2004, respectively. For three years, he had been an assistant professor with the New Mexico Institute of Mining and Technology, Socorro, NM. Currently, he is an Alumni Professor and the director of graduate programs with the Department of Computer Science and Software Engineering, Auburn University. His research interests include parallel and distributed systems, storage systems, fault tolerance, real-time systems, and performance evaluation. Dr. Qin was a recipient of the U.S. National Science Foundation CAREER Award in 2009.

Professor Robert "Glenn" Richey, Jr. (Ph.D. University of Oklahoma) is the Harbert Eminent Scholar and Chair in Supply Chain Management at Auburn University. He is the current EIC or the Journal of Business Logistics. Dr. Richey engages in industry-based supply chain strategy and international business research and enjoys building research and student relationships between academia and practice. Since 2003, Dr. Richey has published more than 90 peer reviewed research articles and taught supply chain and international business courses in 25 countries on six continents. Prior to entering academia, Dr. Richey worked for a decade in managerial positions across several areas including procurement, sales, retail consulting, warehousing, finishing/manufacturing and logistics/supply chain operations.