

## Random Forests

**Deliverable:** All of your work for this assignment should be done in an R Markdown document. Knit your document into a Word file and submit the Word file as the deliverable for this assignment.

**Libraries:** For this assignment you will need the following libraries: tidyverse, caret, and ranger.

Read in the “Blood.csv” dataset. The dataset contains five variables:

Mnths\_Since\_Last: Months since last donation

TotalDonations: Total number of donation

Total\_Donated: Total amount of blood donated

Mnths\_Since\_First: Months since first donation

DonatedMarch: Binary variable representing whether he/she donated blood in March (1 = Yes, 0 = No)

Convert the DonatedMarch variable to a factor and recode the variable so 0 = “No” and 1 = “Yes”.

**Task 1:** Split the dataset into training (70%) and testing (30%) sets. Use set.seed of 1234.

**Task 2:** Create a random forest model on the training set to predict DonatedMarch using all of the variables in the dataset. Use caret’s trainControl function to set up 10 fold cross-validation. Use a random number seed of 123. Use 100 trees (Note you can specify the number of trees by adding a line num.trees = 100 to the rf\_fit block of code).

**Task 3:** Using varImp, what is the most important variable in the model, what is the least important?

**Task 4:** Use the model to develop predictions on the training set. Use the “head” function to display the first six predictions.

**Task 5:** Use the model to create a confusion matrix using caret’s confusionMatrix function for the training set. What is the accuracy, sensitivity, and specificity of the model?

**Task 6:** How does the accuracy of the model compare to a naive model that assumes that all observations are in the majority class?

**Task 7:** Use the model to develop predictions on the test set. Develop a confusion matrix. How does the model perform on the testing set?

**Task 8** Comment on how this model might be used in the “real-world.” Would you recommend this model for real-world use? What if any concerns would you have about using the model?