Project 4 – Clunker Spotting

# Overview

You've been hired by CarVana to look at their used car purchasing process and identify used cars that are likely going to be bad purchases (isBadBuy). CarVana buys cars from people looking to sell their cars but they also purchase used cars at various auctions around the country.

You can read about how to sell your car to CarVana’s here https://www.carvana.com/sell-my-car

And you can read about Manheim auto auctions here <https://www.manheim.com/>.

Anyway, CarVana has a reputation for selling quality used cars, in fact you can buy a car online drive it for 7 days if you don’t like it they’ll pick it up. Your job is to identify the crappy cars before they end up on CarVana’s website for sale. In a nutshell, you are going to build a model to predict how likely a car is going to be a bad buy - i.e the variable isBadBuy. You are required to build and tune two models: a Random Forest and XGBoost model.

You are given two datasets like always

1. Project\_4\_training.csv
2. Project\_4\_ kaggle.csv

What to turn in:

1. **Model evaluation** – keep it simple but well formatted.

* Table of Model Performance comparing your RF and XGBoost on Train and Test, which model performed best, based on your selection metric.
* Table of K-Fold Tuning – you should have at least 5 different experiments for each model using 5 folds so a total of 100 fits.
* ROC chart of Training and Test for RF and XGboost
* Varible Importance for RF and XGboost

1. Your Code
2. Kaggle submission

Kaggle competition link :

<https://www.kaggle.com/t/abd31f49c496401b835e2a53a6afc2da>

Rubirc this project is worth up to 25 points

1. 20 points for your model summary
2. 5 points for your position on Kaggle
   1. 5 points for top 20, 4 for next 20, 3 for next 20, 2 for next 20, 1 for bottom 20% and 0 if you can’t beat the benchmark and -5 if you don’t submit.