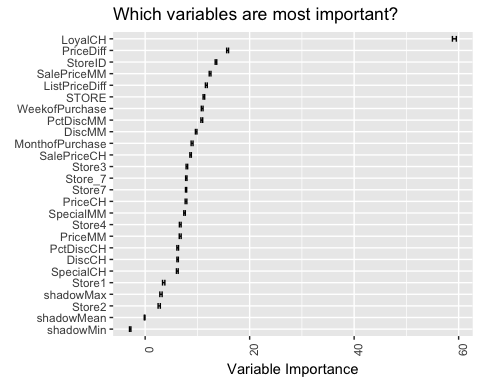
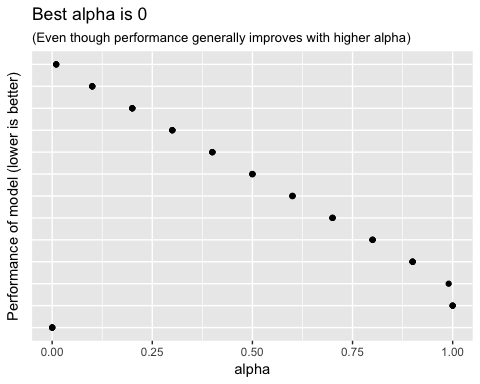
R Notebook





## Confusion Matrix (vertical: actual; across: predicted) for max f1 @ threshold = 0.47984524221999:  
## CitrusHill MinuteMaid Error Rate  
## CitrusHill 128 18 0.123288 =18/146  
## MinuteMaid 21 42 0.333333 =21/63  
## Totals 149 60 0.186603 =39/209

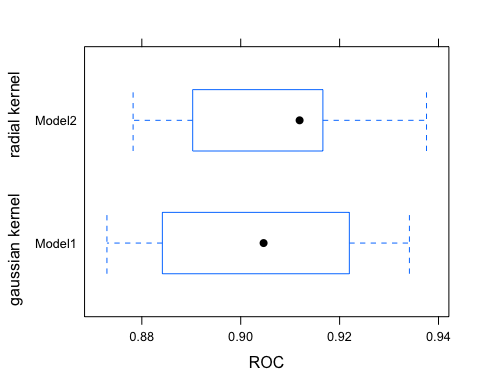
## Confusion Matrix (vertical: actual; across: predicted) for max f1 @ threshold = 0.477382658876859:  
## CitrusHill MinuteMaid Error Rate  
## CitrusHill 128 18 0.123288 =18/146  
## MinuteMaid 20 43 0.317460 =20/63  
## Totals 148 61 0.181818 =38/209

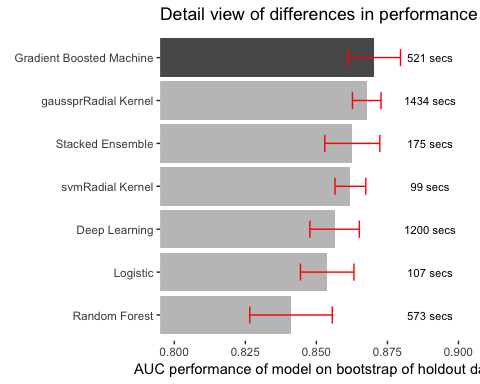
## Confusion Matrix (vertical: actual; across: predicted) for max f1 @ threshold = 0.410677183240095:  
## CitrusHill MinuteMaid Error Rate  
## CitrusHill 414 80 0.161943 =80/494  
## MinuteMaid 51 292 0.148688 =51/343  
## Totals 465 372 0.156511 =131/837

## Confusion Matrix (vertical: actual; across: predicted) for max f1 @ threshold = 0.401450791702452:  
## CitrusHill MinuteMaid Error Rate  
## CitrusHill 414 80 0.161943 =80/494  
## MinuteMaid 51 292 0.148688 =51/343  
## Totals 465 372 0.156511 =131/837

## # A tibble: 5 × 4  
## model ymin y ymax  
## <chr> <dbl> <dbl> <dbl>  
## 1 Logistic 0.8443990 0.8538027 0.8632064  
## 2 Deep Learning 0.8477040 0.8564074 0.8651107  
## 3 Random Forest 0.8265549 0.8410691 0.8555832  
## 4 Gradient Boosted Machine 0.8611365 0.8703672 0.8795980  
## 5 Stacked Ensemble 0.8529524 0.8626322 0.8723120

##   
## Call:  
## summary.resamples(object = .)  
##   
## Models: Model1, Model2   
## Number of resamples: 10   
##   
## ROC   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## Model1 0.8729 0.8887 0.9046 0.9045 0.9198 0.9341 0  
## Model2 0.8782 0.8928 0.9119 0.9066 0.9156 0.9376 0  
##   
## Sens   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## Model1 0.7143 0.7647 0.7647 0.7756 0.7697 0.8824 0  
## Model2 0.7353 0.7647 0.7828 0.7842 0.7985 0.8824 0  
##   
## Spec   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## Model1 0.8163 0.8644 0.8788 0.8865 0.9145 0.94 0  
## Model2 0.7755 0.8571 0.8776 0.8743 0.9138 0.92 0



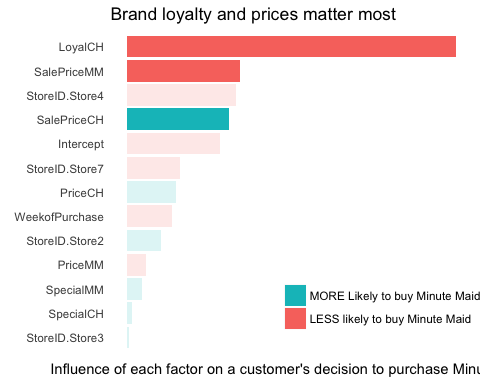


# Overview

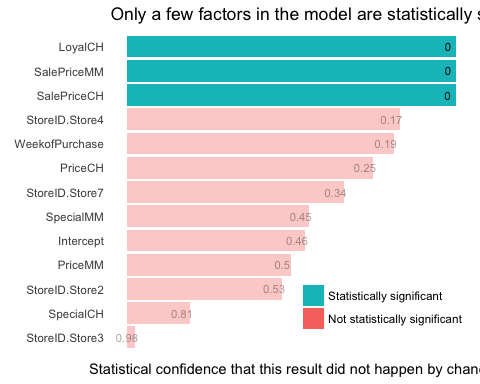
The goal of this analysis is to understand how to make the Orange Juice category of store sales perform better. The store sells two brands of orange juice, Minute Maid (MM) and Citrus Hill (CH). Since MM has higher margins than CH, this analysis will make recommendations regarding which factors influence a consumer's decision to purchase MM orange juice. This allows our compaqny to leverage those factors as opportunities to influence MM sales. It will additionally provide a predictive model for more precise forecasting. This forecasting will be of benefit now, but will be of tremendous benefit later when the company adjusts its marketing to increase MM sales (since an updated forecast will be required).

# Problem Definition

Brand manager 1. What predictor variables influence the sales of MM?



1. Are all the variables in the dataset effective or are some more effective than others?



1. How confident are you in your recommendations?

