­­

Project 1

An in-depth analysis of Orange Juice sales

Dan Lemaire - Gavin Moss - Joni Jensen

Submission Date: June 1st, 2017

Mktg6620 – Machine Learning for Business Applications

Overview 2

Problem Definition 2

Executive Summary 2

Drivers 3

Predictive Model 5

Methods 5

Data Prep 5

Model Selection 5

Variable importance 5

Recommendations 6

Loyalty 6

Data 6

Pricing 6

Conclusion 6

# Overview

The goal of this analysis is to understand how to increase the revenue from the orange juice category of store. 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 company 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

The stores sell two brands of orange juice, Citrus Hill and Minute Maid. As Minute Maid sales have a lager profit margin we want to find what factors cause customers to purchases Citrus Hill orange juice instead of Minute Maid. We are also interested in building a model that is able to predict with confidence what juice a customer will purchase.

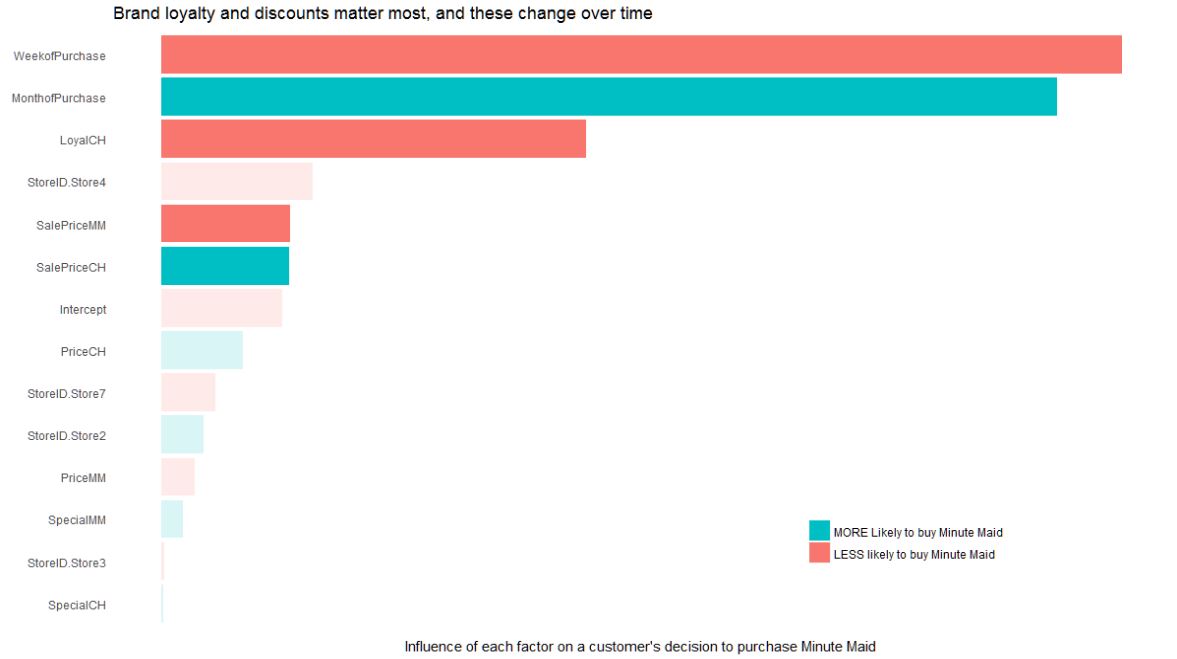
## Executive Summary

Through an analysis of a general linear model we have three suggestions on how to increase the sales of minute maid, and therefore increase the revenue from orange juice sales. Assuming that all buyers will continue to buy one of the two brands carried.

* Erode the Loyalty customers feel for Citrus Hill orange juice.
* Raise the price of Citrus Hill.
* Check for Seasonality.

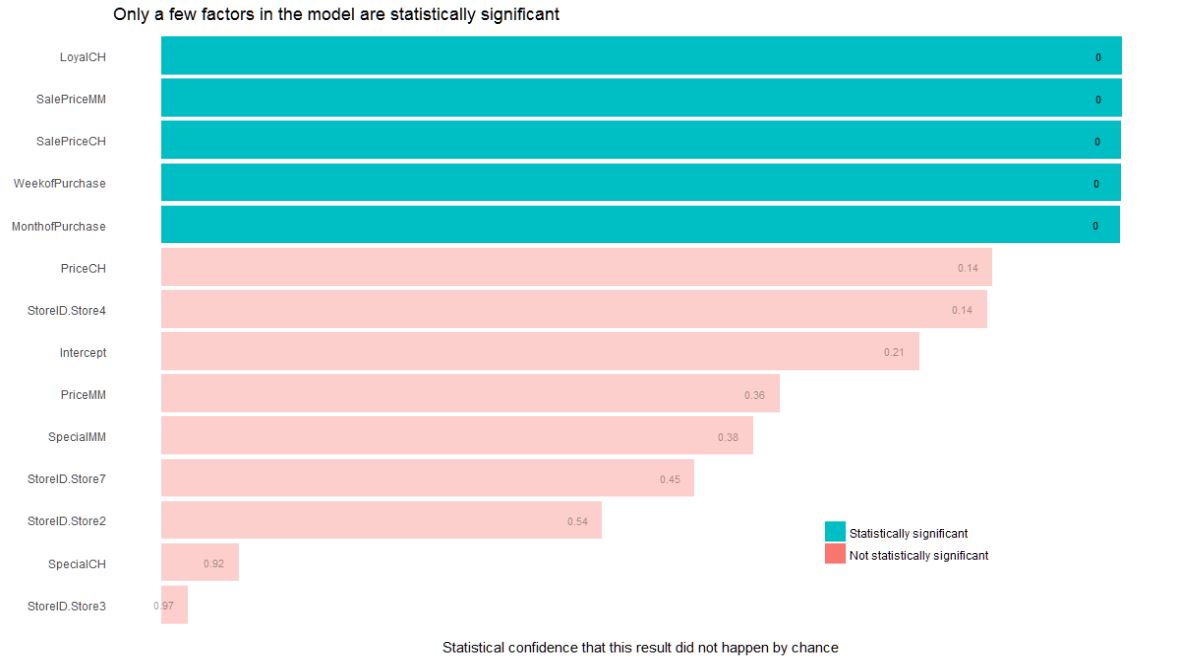
# Drivers

The following chart shows the magnitude of the effect different drivers have on a customer’s propensity to buy Minute Maid and Citrus Hill orange juice.



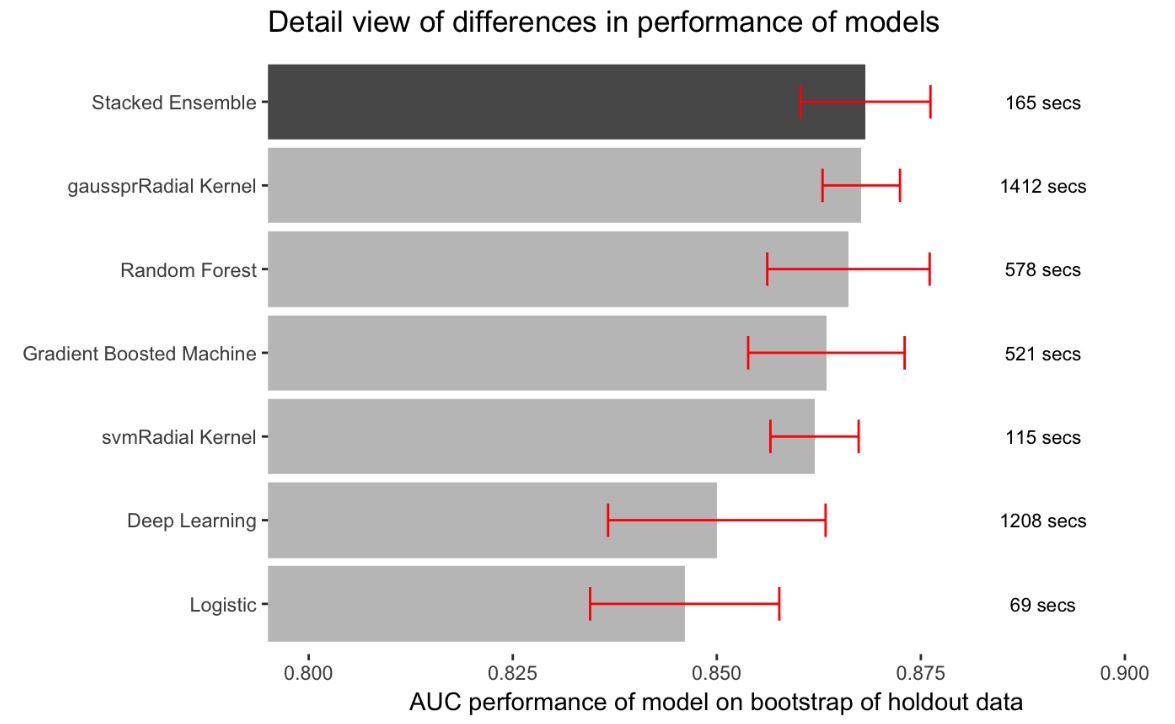
We see that the weekofPurchase has a large negative effect on a customer buying Minute Maid but the MonthofPurchase has a large positive effect. It is unclear here as to why this is the case without more data. We hypothesis that there is some amount of seasonality that effects orange juice sales but that cannot be tested with only one year of data.

Other prominent drivers are if the customer shopped at store 4, the Loyalty a customer has for Citrus Hill, and the sale price of both minute Maid and Citrus Hill. However the effect of store 4 is statistically insignificant, as showing in the chart below.



# Predictive Model

For the purpose of prediction we evaluated eight models. The chart below shows the AUC performance scores for each model. All the models evaluates performed well with scores between .845 and .865. Our best model was a stacked ensemble of a general leaner model, a gradient boosting model, and a deep learning model. Which Preformed marginally better than a support vector machine using a gaussprRadial Kernel. It is worth noting that the SVM model preformed more consistently then the ensemble model, but it also took over eight times as long to run. In the end either one of these models could be used for predicting if customer will buy Minute Maid or Citrus Hill. The choice should come down to how the model will be implemented and if the run time is a concern.



# Methods

## Data Prep

## Model Selection

## Variable importance

# Recommendations

## Loyalty

## Data

## Pricing

# Conclusion