**R-Shiny Decision Support System**

**AVOCADABRA – Forecasting Avodaco Demand**

**1.      Business Problem Definition:**

With the avocado demand and prices at an all-time high in the US, retailers, wholesalers and stakeholders in the avocado business strive to achieve optimal inventory management. The stakeholders seek ways to minimize inventory, minimize stock-out, and meet the demand.

Since this required identifying patterns in historical data, and drawing predictions from the same, it was formulated as an analytical problem. This allowed forecasting future demand using analytical methods.

The stakeholders of this app were identified as retailers and wholesalers of avocado. Predicting future demand would allow the stakeholders to efficiently manage their inventory and enhance their pricing strategies.

**2.      Analytics Problem Definition:**

The business problem was then reconstructed as an analytical one as follows.

With the avocado demand and prices at an all-time high in the US, retailers, wholesalers and stakeholders in the avocado business strive to achieve optimal inventory management. Based on historical data, forecast demand and predict order quantities that would efficiently manage inventory and strategize pricing.

Drivers identified were region of the avocado market, date being predicted for, price of avocado, type of avocado (conventional or organic), seasonality pattern in the demand of avocado which was reflected in quantities sold each month.

It is assumed that the analysis is limited to data provided by the Hass Avocado Board website. The analysis does not consider external factors like disaster index which can influence the supply of avocados.

The key metrics of defining success would be if the user adhered to the predicted numbers and does not face events of overstocking or understocking.

**3.      Data**

The dataset ‘Avocado Prices’ was selected from [www.kaggle.com](http://www.kaggle.com). The data was clean with no missing values. There were 18249 observations spread over a period over four years. This time span was enough to build forecasting model. The columns type, regions, year were converted into factors. Columns total bags and total volume columns were dropped. The date column was formatted to the date format.

It was expected that the average sales and average price of avocados would have a dependency on the month due to seasonality.

**4.      Methodology Selection:**

The following analytical approaches were used. Generalized Linear Regression on the data gave a low R-sq value after which adaptive lasso was used. This still did not yield a large enough R-sq value. This could be because time series models do not follow linear patterns. Hence Auto ML was employed to give the most optimal model, and this gave gradient boosting with an R-sq of 0.98.

R programming language is most compatible with Shiny apps. Shiny apps have an easy-to-use interface and can be used by all kinds of users.

**5.      Model Building:**

Calibration: We first divided the data into three sets: train, validate and test. We used the 2015 and 2016 data in training frame; 2017 data in validation frame and 2018 data in test frame. We then passed the train data to train potential models in auto ML using h2o package. The leader model came out to be GBM (gradient boosting model), which gave an adjusted R squared value of 98%.  This model has 115 number of trees and max\_depth of 7 as parameters, among others. Thus, we could ensure that the model generalizes well with the data. We integrated this model in the R Shiny app to predict the volume of avocados (dependant variable)

When the user selects the date, the earliest date that can be selected is one day after the system date. Once the app predicts the order quantity, the trend of units ordered in the past is also displayed as a time series graph.

Exploratory Data Analysis was carried out as follows using ggplot2. A function was created which takes region and price as the input variables and gives appropriate EDA plots:

1. Trend of average monthly sales over the years for the selected regions (for types organic and conventional)
2. Trend of monthly average prices over 4 years for that particular region

**6.      Functionality:**

The DSS that was developed can predict the future demand of avocados for a particular date, based on input parameters such as region, the average price of an avocado at the time of presentation, and the type and subcategory of avocado. The DSS also plots historical data (exploratory data analysis) which provides price for each month grouped by avocado type (conventional or organic) and sales based on regions. Similarly, there is an option to plot sales for each month. These trendlines help the user identify seasonality in the markets. The packages we used and found useful for data cleaning, manipulation and analysis were:

1. lubridate: date handling,
2. h2o,
3. dplyr – manipulating data,
4. Ggplot2- plotting graphs and visualization,
5. scales: time series plot,
6. grid: multiplot
7. Additional packages used: Shiny, shinythemes

Given some additional time, we could improve the aesthetics of the app. We could also include more UI functionality within the app.

**References :**

1. Login Page : <https://stackoverflow.com/questions/32026079/how-to-adjust-the-position-of-conditional-panel-in-shiny>
2. Time-series graph and Model: <https://rpubs.com/phamdinhkhanh/390053>
3. Shiny App Design Ideas : <https://github.com/liu1498/Movie-Recommender-System-ShinyApp>
4. Multiplot function used in class work has also been re-used in our app.
5. Business Understanding : <https://hassavocadoboard.com/>
6. Dataset : <https://www.kaggle.com/neuromusic/avocado-prices>