Week 13 Deliverables

Group Name: Data Forecasting Team

Team Member's Details

Name: Charles Ibitamuno

Email: charlesibitamuno@gmail.com

Country: United States of America

College/Company: Columbia University of General Studies

Specialization: Data Science: Retail Forecasting

Problem Description:

The large beverage company in Australia needs to forecast demand for each of their products at the item level, on a weekly basis. Their sales are influenced by various factors including promotions, holidays, and seasonality. The company currently uses an in-house software solution for forecasting, but it often produces unreliable results. They want to explore AI/ML-based forecasting to replace their current system.

Final Project Report:

From what I've observed in these past few weeks from the data in the forecasting case study, I noticed that the sales of these products would increase whenever there was an In-Store or End Promotion, which would explain the significant spike in sales throughout the years. I also observed that during the COVID less people bought these goods, foreshadowing the negative effects a pandemic has on such beverages. Ironically, the festive seasons weren't the cause of these spikes; showing us that despite what it may seem, seasonal holidays don't improve the sales of these drinks.

The Price Discount % feature does play a positive role in these sales, as long as the range is no more than 58%. After that, there's a slight increase before it drops, informing us that one mustn't drop the price below 50% if they want people to be comfortable to buy them.

The Google Mobility feature didn't do much to move the product sales. Even though the Correlation matrix has a higher value than the Seasonal holidays, it doesn't influence the spike pattern we're interested in.

In developing a model to train the sales forecast, I came up with four different models: one that is linear (*Linear Regression Model*) another for boosting (*Gradient Boosting Regressor*) an Ensemble Model (*Random Forest Regressor*), and the *Support Vector Regressor(SVR)* to help predict continuous values. Below are the accuracy results for each model:

Linear Regression - Accuracy: 0.3285, Execution Time: 0.00 seconds

Random Forest - Accuracy: 0.5461, Execution Time: 0.34 seconds

SVR - Accuracy: 0.3327, Execution Time: 0.10 seconds

Gradient Boosting - Accuracy: 0.5296, Execution Time: 0.10 seconds

Best model: Random Forest with accuracy 0.5461

Models sorted by accuracy:

Random Forest: Accuracy = 0.5461, Execution Time = 0.34 seconds

Gradient Boosting: Accuracy = 0.5296, Execution Time = 0.10 seconds

SVR: Accuracy = 0.3327, Execution Time = 0.10 seconds

Linear Regression: Accuracy = 0.3285, Execution Time = 0.00 seconds

With this conclusion, it is safe to say that the accuracy of the Random Forest Regressor is better suited for this forecast analysis, giving the marketers a better idea of how their sales would look in the future.