Project Progress Report

INFO 7390 : Advances in Data Sciences and Architecture

Grocery Sales Forecasting

Submitted by:

Chaitanya Joshi (001860018)

Mayank Gangrade (001837520)

Rohan Naik (001824278)

1. Abstract

Product sales forecasting is a major aspect of purchasing management. Forecasts are crucial in determining inventory stock levels, and accurately estimating future demand for goods has been an ongoing challenge, especially in the Supermarkets and Grocery Stores industry. If goods are not readily available or goods availability is more than demand overall profit can be compromised. As a result, sales forecasting for goods can be significant to ensure loss is minimized. Additionally, the problem becomes more complex as retailers add new locations with unique needs, new products, ever transitioning seasonal tastes, and unpredictable product marketing. In this analysis, a forecasting model is developed using machine learning algorithms to improve the accurately forecasts product sales. The proposed model is especially targeted to support the future purchase and more accurate forecasts product sales and is not intended to change current subjective forecasting methods. A model based on a real grocery store's data is developed in order to validate the use of the various machine learning algorithms. In the case study, multiple regression methods are compared. The methods impact on forecast product availability in store to ensure they have just enough products at right time.

2. Introduction

In this project, we are trying to forecasts product sales based on the items, stores, transaction and other dependent variables like holidays and oil prices.

This is a Kaggle Competition called "Corporación Favorita Grocery Sales Forecasting" where the task is to predict stocking of products to better ensure grocery stores please customers by having just enough of the right products at the right time.

For this particular problem, we have analyzed the data as a supervised learning problem. In order to forecasts the sales we have compared different regression models like Linear Regression, Decision Tree, ExtraTreeRegressor, Gradient Boosting, Random Forest and XgBoost. Further to optimize the results we have used multilayer perception (MLP: a class of feed forward artificial neural network) and LightGBM (gradient boosting framework that uses tree based learning algorithms).
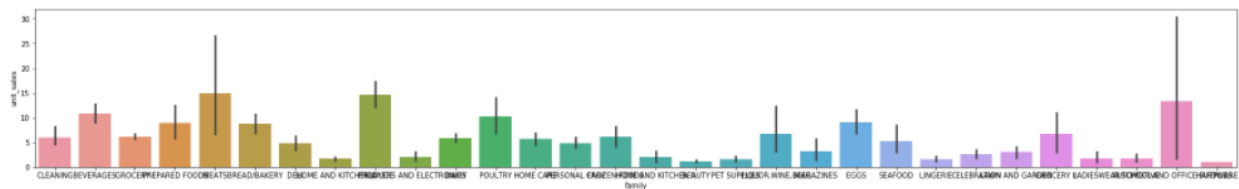
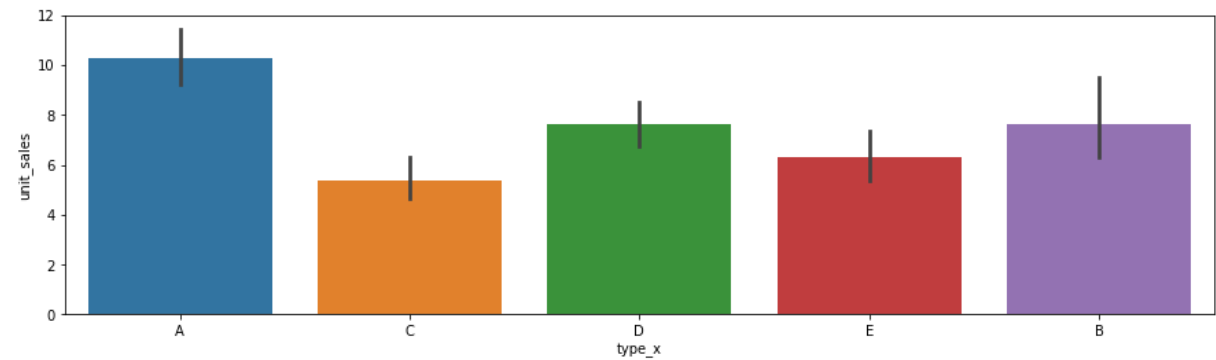**3. Code documentation**

Refer to the following link :

https://github.com/mgangrade7/INFO-7390-A`DS-Project.git
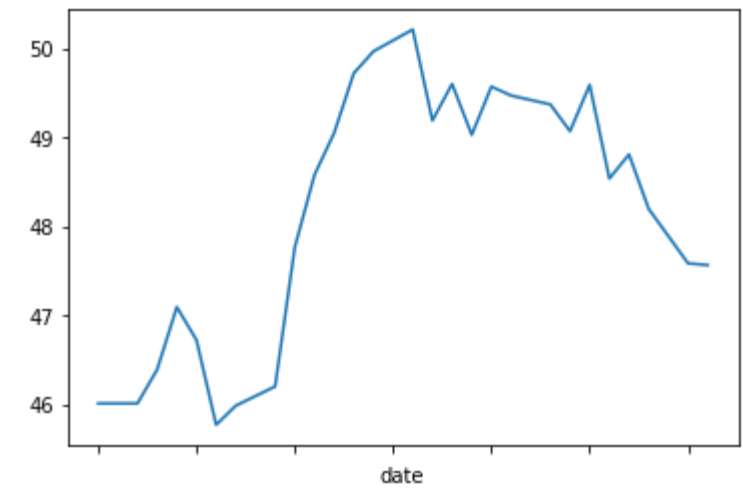
4. Observation and Results:

Sales per Item Family

Sales per Store Type

Oil price changes

We are getting following results on applying data set on different models:

Model	R2 Score
Linear Regression	0.354
Decision Tree Regression	0.705
Extra Tree Regression	0.825
Random Forest Regression	0.836
Gradient Boosting Regression	0.836
XG Boost	0.797
LGBM	0.759

5. Discussion

Sales Forecasting is the process of using the company's sales records of the past years to predict the short-term or long-term performance in the future. This is one of the pillars of proper financial planning. As with any prediction-related process, risk and uncertainty are unavoidable in Sales Forecasting too. Hence, it's considered good practice for forecasting teams to mention the degree of uncertainties in their forecast.

Accurately forecasting sales and building a sales plan can help to avoid unforeseen cash flow problems and manage production, staff and financing needs more effectively.

6. References

[1] Cui, G., Wong, M. L., & Lui, H. K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming.Management Science, 52(4), 597-612

[2] Taylor, E. L. (2014). Predicting Consumer Behavior. Research World, 2014(46), 67-68

[3] Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales?. International Journal of Forecasting, 23(3), 347-364

[4] https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data

[5] https://en.wikipedia.org/wiki/Xgboost

[6] https://en.wikipedia.org/wiki/Random_forest

[7] https://en.wikipedia.org/wiki/Decision_tree

[8]https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/

[9] https://www.tutorialspoint.com/sales_forecasting/sales_forecasting_discussion.htm