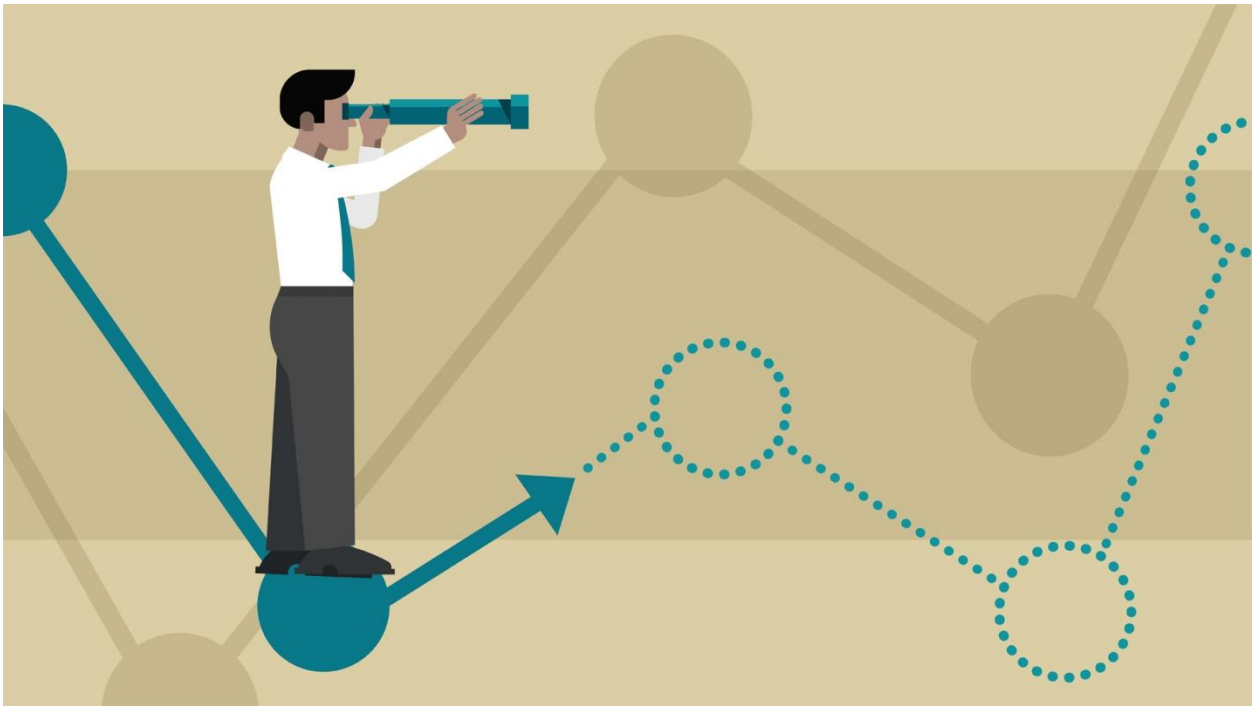
Grocery Sales Forecasting

Seelam Venkata Krishna Reddy

15/10/2022

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

1. Problem Statement

The problem statement 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

2. Market/Customer/Business need Assessment

The aim of this project is to forecast more accurate product sales for the Ecuadorian supermarket chain based on certain features.

3. Target Specification

The proposed system/service will provide the shopkeepers and vendors with some techniques so that their sales boost up and they no longer have to go through an economic crisis.

4. External Search

The sources I have used as reference for analyzing the need of such a system

[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-vsxgboost/

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

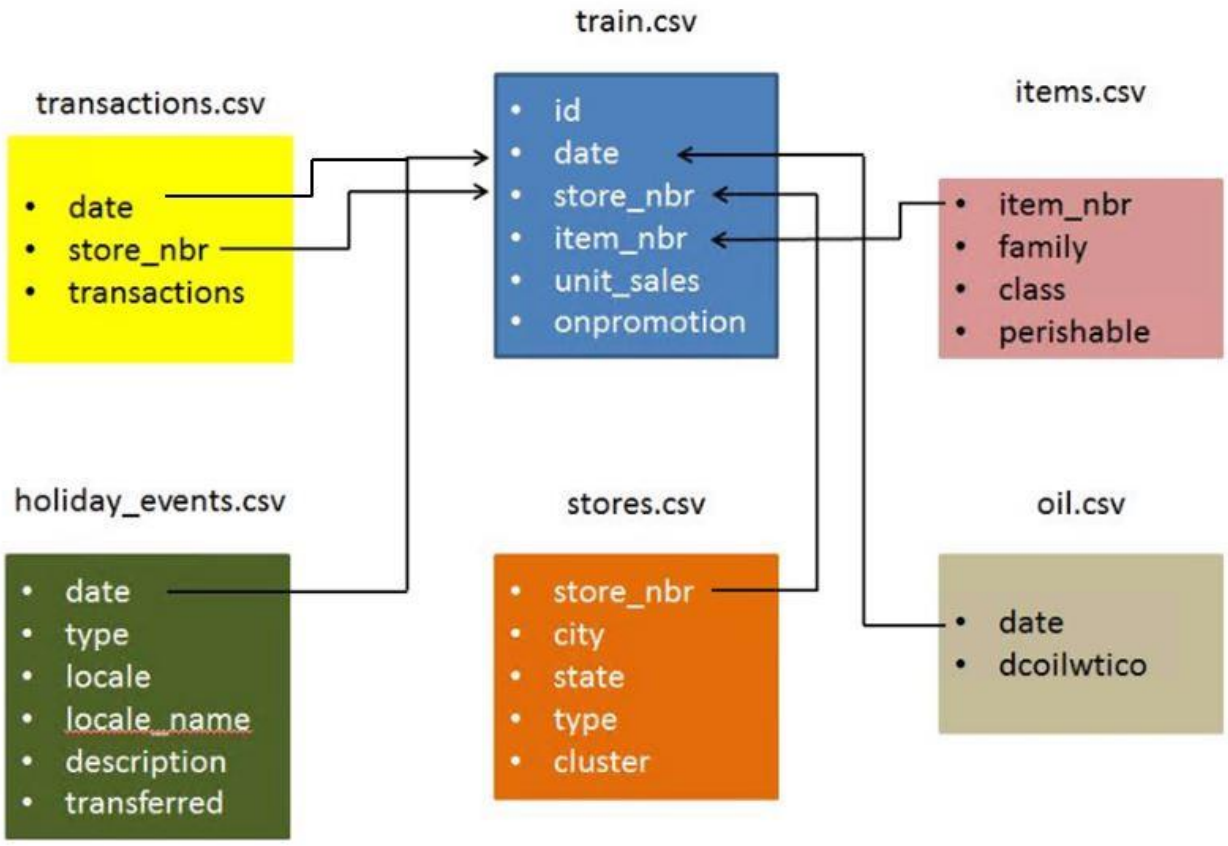
5.Applicable Constraints

- Data Collection from shopkeepers and vendors
- Continuous data collection and maintenance
- Convincing the shopkeepers to implement the system in their shops.
- Taking care of rarely bought products
- Lack of technical knowledge for the user(vendors)

6. Applicable Regulations

- Data protection and privacy regulations(Customers)
- Govt Regulations for small businesses
- Employment Laws
- Antitrust Regulations
- Regulations against false advertising

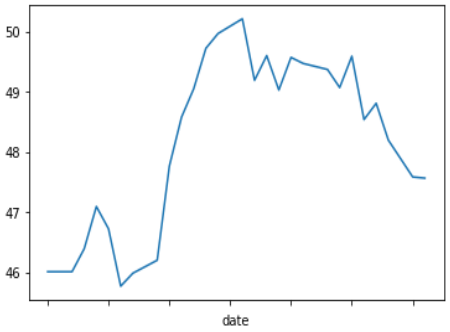
8.Final Product Prototype



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).

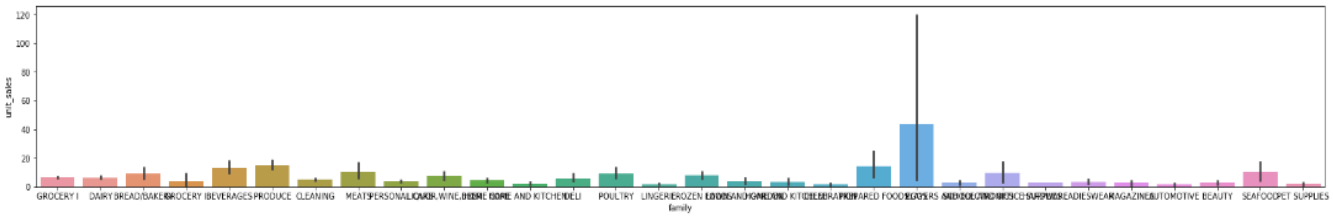
9.Code implementation

```
#Plot the oil values
oil_plot = oil['dcoilwtico'].copy()
oil_plot.index = oil['date'].copy()
oil_plot.plot()
plt.show()
```



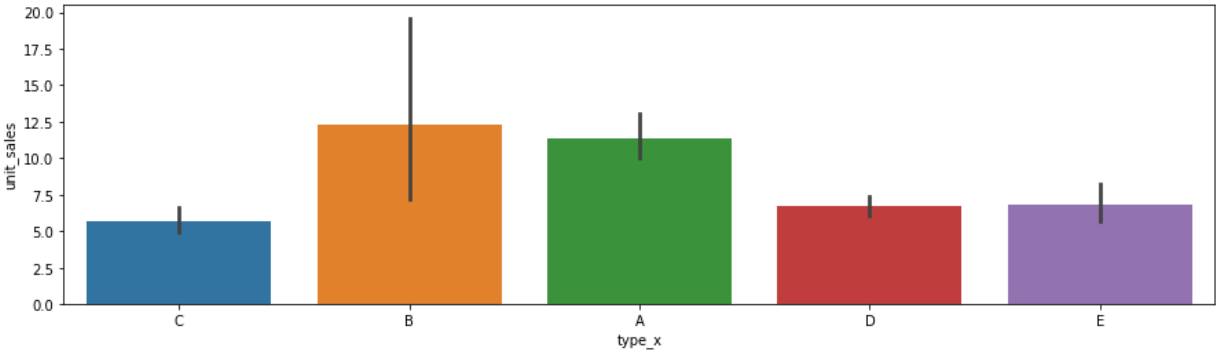

```
# Plotting Sales per Item Family
fig, (axis1) = plt.subplots(1,1,figsize=(30,4))
sns.barplot(x='family', y='unit_sales', data=strain, ax=axis1)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py:1460: FutureWarning: remove_na is deprecated and is a private function. Do not use.
  stat_data = remove_na(group_data)
<matplotlib.axes._subplots.AxesSubplot at 0x1d780cfeb00>
```



```
# Plotting Sales per Store Type
fig, (axis1) = plt.subplots(1,1,figsize=(15,4))
sns.barplot(x='type_x', y='unit_sales', data=strain, ax=axis1)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\categorical.py:1460: FutureWarning: remove_na is deprecated and is a private f
  stat_data = remove_na(group_data)
<matplotlib.axes._subplots.AxesSubplot at 0x1d780d2ca20>
```



10.Conclusion

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.