

Product Sales Predictor

Abdul Wahid
Software Engineering
wahidiccq@gmail.com

Abstract— There are many retail stores providing different type of selling of merchandise and certain services to consumers. Some of the retail stores provide their services globally in different countries. Those businesses are facing challenges due to unexpected demands and run out of stock some times due to the inappropriate machine learning algorithm or some times if you have got high levels of excess stock, the chances are you have low inventory turnover, which means you are not turning all your stock on a regular basis and unfortunately, excess stock that sits on warehouse shelves begin to expire. Moreover, you need space for your products to store and this can tolerate costs to maintain. An ideal ML algorithm will predict demand accurately and algorithm used in machine learning facilitates for retailers to have personalized product recommendations for different customers based on customer's purchase trends and unique interest. Besides improving customers experience, machine learning can help retailers to section their products in their stores accurately. This project can be used by almost every retail store across the globe to keep track of the stock of their products. Using this app will help the storeowners to predict the sales of products in any specific time and then if more stock is needed, then they can order more according to the demand in the current market. This app just has to know all the data about previous sales of the products in last few years in specific times and then it will automatically generate the sales of the products in the coming time. In this way, only that mount of stock can be ordered for the store which will neither be short at the store, nor will it be in great amount so the chances of it being expired can be reduced.

Keywords—

I. INTRODUCTION

One of the most common financial decisions that each of us makes on a nearly daily basis involves the of various products, goods and services. Data is growing in massive amount on internet and time plays very important role in every person's life. It is impossible for a single person to read whole data daily. Sometimes decision regarding whether or not to make a purchase is dependent on price but in many cases the purchasing decision is more complex. Retailers nowadays understand this well and attempt to make use of it in an effort to gain an edge in a highly competitive market. This is

specially done in an effort to make purchasing more likely, in addition to balancing the scalability and profit in setting the selling price of a product. Companies frequently introduce additional elements to the offer which are aimed at increasing the perceived value of the purchase to the customer. An important aspect of managing supply chain efficiently is to have better prediction of sales such that manufacturer will not over or under purchase production products. An emerging area in prediction of sales is in big data and user-generated content on the sales of product. Given that user-generated content plays an important role in influencing the purchasing decisions of consumers, its role in helping organizations to understand and predict product demand can be investigated. User generated data is nothing but the data which is generated but users itself when he/she gives ratings, reviews etc in a particular website.

VARIABLES	DESCRIPTION
Discount Value	The monetary value of price deduction from usual price.
Discount Rate	The percentage of price deduction
Current Price	The price of the product
Free Delivery	Whether the product is delivered without delivery fee
Customer Review Rating	The accumulated average numeric rating of in line reviews
Number of customer Reviews	The number of all online reviews
Percentage of Negative Review	The proportion of 3-star reviews in total reviews
Percentage of Positive Review	The proportion of 4-star and 5-star reviews in total reviews

II. LITEARATURE REVIW

This approach was proposed by Deven Ketkar. In this methodology raw data collected at big mart was pre-processed for missing anomalies and outliers. Then an algorithm was trained on this data to create a model .Algorithms used were Random forests and multiple Linear Regression. ETL that is Extract, Transform and Load tool was used in this methodology to get data from one database and transform it into suitable format. Data was transformed from sample raw data into understandable format. The model was used for final results. This approach was proposed by Bohdan M. Pavlyshenko. This

methodology is a stacking approach for building regression. Ensemble of single models was studied for implementation. The algorithms used were Random forest and regression. The results showed that using stacking techniques we can improve performance of predicting model. Sales was based on old dataset and not on usergenerated data. In earlier projects ,were stored dataset was used the prediction was not that accurate. Dataset which was used in the project was 2-3 years old and the sales which it was predicting now was on the basis of that data. Nowadays data is being generate data such a large rate and there would be so much changes in new data in comparison with the stored data

III. METHODOLOGY

Web scraping will be done using a web crawler. Wrapper program would be used to detect templates in source. Required real time data is gathered and copied from the web and stored in a file for process. B. Classification

Algorithm used for Classification:

Natural Language processing algorithms

-It is concerned with the interactions between computers and human languages.

-Its main objective is to read, understand and make sense of human language in a manner that is valuable.

Natural Language Toolkit (NLTK) provides libraries for classification.

Parameter that algo uses:

Customer Reviews .

C. Prediction

Algorithm used for Prediction:

Multiple Linear Regression

-It is a statistical technique that uses various explanatory variables to predict the outcome of response variable.

Formula is $y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n$

Where y=dependent variable and x=independent variables

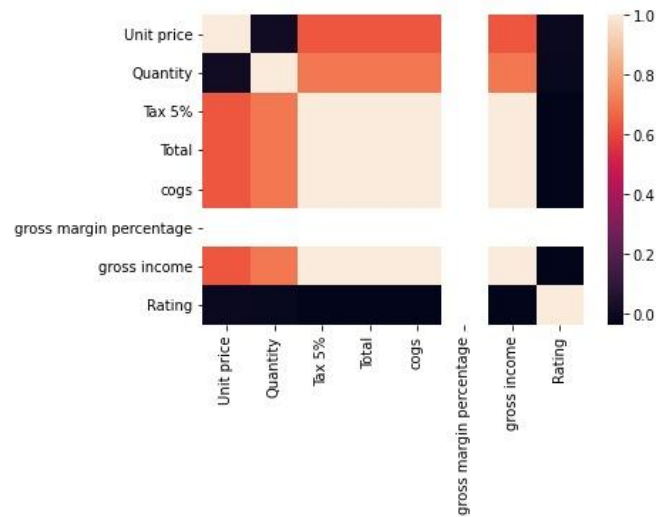
Parameters that algo uses:

- 1.Sentimental analysis of Reviews
- 2.Online review Volume

IV. RESULTS AND CONCLUSION

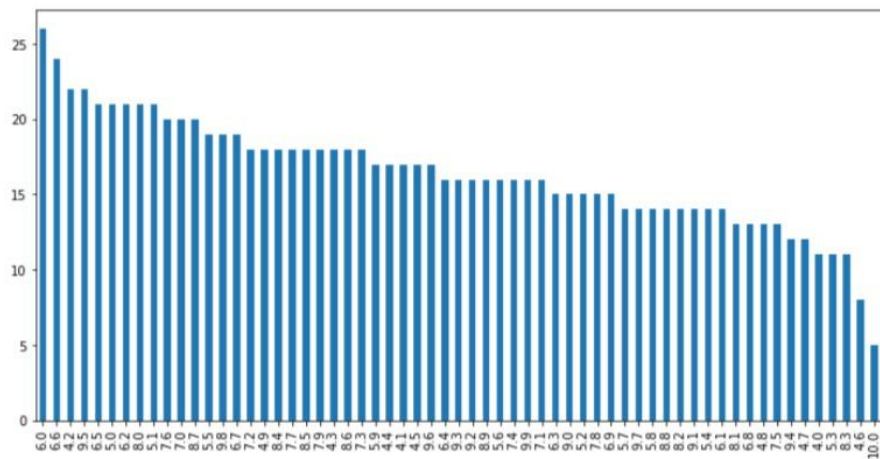
```
#Checking correlation between columns within the dataset  
sns.heatmap(df.corr())
```

<AxesSubplot:>



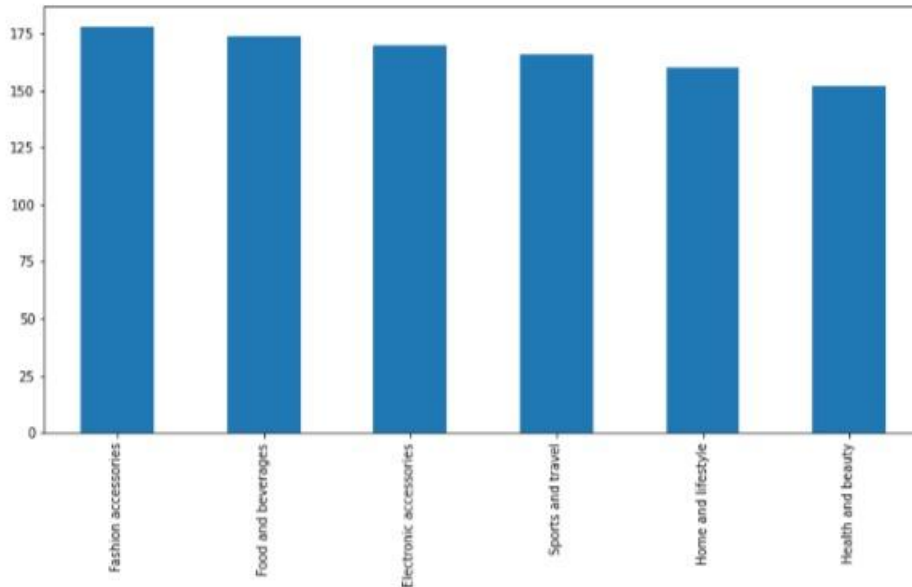
```
plt.figure(figsize = (12,6))  
df['Rating'].value_counts().plot(kind="bar")
```

<AxesSubplot:>



```
plt.figure(figsize = (12,6))
df['Product line'].value_counts().plot(kind="bar")
```

<AxesSubplot:>



V. CONCLUSION

Market basket analysis generates the frequent item set i.e. association rules can easily tell the customer buying behaviour and the retailer with the help of these concepts can easily setup his retail shop and can develop the business in future. The main algorithm used in market basket analysis is the Linear Regression algorithm. It can be a very powerful tool for analyzing the purchasing patterns of consumers. The three statistical measures in market basket analysis are support, confidence. Support measures the frequency an item appears in a given transnational data set, confidence measures the algorithm's predictive power or accuracy. In our example, we examined the transnational patterns of grocery purchases and discovered both obvious and not-so-obvious patterns in certain transactions. Association rules and the existing data mining algorithms usage for market basket analysis, also it clearly mentioned about the existing algorithm and its implementation clearly and also about its problems and solutions. Predictive modelling offers the potential for firms to be proactive instead of receptive. Predictive modelling using transnational data create particular challenges which need to be carefully addressed to develop valuable models. Consumers, planners, merchandisers and store administrators have started to recognize how this new era of easy-to-use market basket analysis tools helps to work more

intelligent and compete more successfully. Our future work would be to design and develop intelligent prediction models to generate the association rules that can be adopted on recommendation system to make the functionally more operational. Better and effective rule mining techniques can be used for better performance of the recommendation system.

References:

Mr. Soham Patangia
B.E Student, Electronics and Telecommunication
Rajiv Gandhi Institute Of Technology,
Mumbai, India

Simsri Panchal², Tanvi Sawant³, Janhavi Shinde⁴
U.G Student,
Dept. of Computer Engineering,
Bharati Vidyapeeth College of Engineering,

https://www.researchgate.net/publication/341870759_Predicting_Online_Product_Sales_using_Machine_Learning

https://www.researchgate.net/publication/345897383_Sales_Prediction_of_Market_using_Machine_Learning