Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Predicting Online Product Sales using Machine Learning

Sandeep Chavan[1]
Assistant Professor,
Dept. of Computer Engineering,
Bharati Vidyapeeth College of Engineering,
Mumbai, India

Simsri Panchal[2] ,Tanvi Sawant[3], Janhavi Shinde[4]
U.G Student,
Dept. of Computer Engineering,
Bharati Vidyapeeth College of Engineering,
Mumbai, India

***Abstract*— Product sales prediction is a major aspect of purchasing management. One of the key challenges faced nowadays by organizations the dynamic, international and unpredictable business environment in which they operate. With growing customer expectations for price and quality, manufacturers today can no longer rely only on cost advantage that they have over their rivals. Forecasting the sales are crucial in determining inventory stock levels and accurately estimating the future demand for goods has been an ongoing challenge in industries If goods are not readily available or if goods availability is more than demand overall profit can be compromised. As a result, sales prediction for goods can be significant to ensure that loss is minimized.Depending on this study, our project is creating a prediction model using machine learning algorithms for accurately predicting online product sales. Our project aims to use uptodate data which includes online reviews ,online ratings ,online promotional strategies and sentiments and various other parameters for predicting product sales.**

Keywords—SalesPrediction, Online products, Machine Learning

I. INTRODUCTION

One of the most common financial decisions that each of us makes on a nearly daily basis involves the purchasing 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. This user generated data plays a very important role in sales analysis in the ecommerce industry.

II. LITERATURE SURVEY

A.FORECAST FOR BIG MART SALES

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.

B.SALES TIME SERIES FORECASTING

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.

III. ANALYSIS OF STUDIED SYSTEMS

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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

More samples did not improve the accuracy

In some systems Random forest algorithm is used.Random forest gives accurate prediction for small datasets.But if the project is using larger dataset the accuracy does not increase on increasing the dataset.

Appropriate parameters were not considered

There is certain correlation between the parameters which affect the sales. The parameters who have better correlation amongst them must be considered for more accurate prediction.

IV. PROPOSED SYSTEM

The study helped to design a model which can facilitate future business researches for predicting product sales in an online environment. The main objective of the project is to show that product demands can be predicted through the comparative influence of promotional marketing strategies such as discounts and the provision of free delivery choices, user generated contents such as volume and valence of on-line reviews ,and sentiments of the web reviews. The algorithms use asynchronous I/O (input/output) to request, extract and pre-process data in real-time from Amazon.com using a web crawler. After getting the data, the texts of reviews are then processed using natural language processing (NLP) algorithm .The resulted sentiment is labelled as positive, negative or neutral for further analysis. This study will then use a Multiple Linear Regression to predict product sales, as well as to predict the effects of the online sentiments on the same so as to design effective promotional strategies and sales tactics.

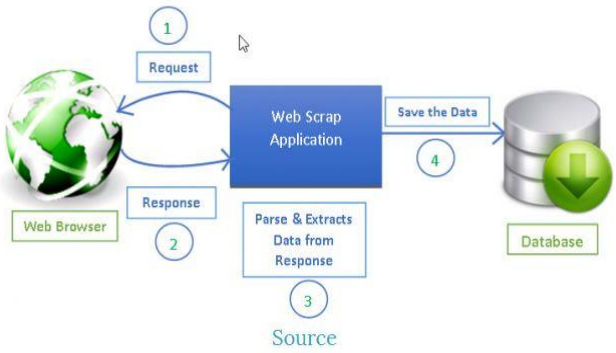
Parameters considered for prediction.

VARIABLES	DESCRIPTION
Discount Value	The monetary value of price deduction from usual price
Discount Rate	The percentage of price deduction from usual price
Current Price	The price of the product
Free Delivery	Whether the product is delivered without delivery fee
Customer Review Rating (Valence)	The accumulated average numeric rating of inline review
Number of customer Reviews (Volume)	The number of all online reviews
Percentage of Negative Review	The proportion of 1-star and 2-star reviews in total reviews
Percentage of Neutral 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
Review text Sentiment (Sentiment)	The sentiment of most helpful reviews
Number of Answered Questions	The number of answered questions in Customer Question & Answers
Manufacturer	The name of the manufacturer of the product
Sales Rank	The Best Sellers Rank of the product

V . METHODOLOGY

A. *Extraction*

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=b0+b1*x1+b2*x2+.....bn*xn**
Where y=dependent variable and x=independent variables
Parameters that algo uses:

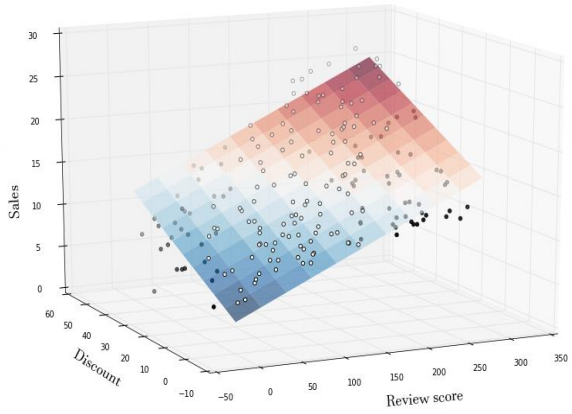
1.Sentimental analysis of Reviews
2.Online review Volume

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

3.Free Delivery
4.No. of customer Q. answered
5.Discount Value
6.Online Ratings

V. CONCLUSION

The paper proposes that promotional marketing strategies and social interactions such as online review and answered questions are both important for influencing sales. The paper shows that sentiments has a significant interaction with volume and valence of online review and could significantly affect and predict product sales. In summary, we have shown that when sentiments interacts with volume and valence, it becomes a more important predictor of product sales.



VI. REFERENCES

[1] East, R.,Hammond, K. and Lomax,W.(2008), "Measuring the impact of positive and negative word of mount on brand purchase probability ",International Journal of research in Marketing ,Vol.25 No. 3,pp.215-224

[2] Cui, G., Lui, H. and Guo, X.(2012), "The effect of online consumer reviews on new product sales", International Journal of Electronics , available at:http://www.tandfonline.com/doi/abs/10.2753/JEC1086-4415170102(accessed 10 March 2015)

[3] Bohdan M. Pavlyshenko (2018)," Machine Learning models for sales time series forecasting"

[4] Professor Deven Ketkar(2018). "A Forecast for Big Data Sales based on Random Forests and Multiple Linear Regression", IJEDR 2018, VOL. 6, ISSUE 4. ISSN: 2321-9939