ISSN: 2584-0495 Vol. 1, Issue 6, pp.418-422

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Design of a prediction system for anticipating the consumer's purchase intention of durable goods

Shaik Qadeer, Mohammed Yousuf Khan

Cite as: Qadeer, S., & Khan, M. Y. (2023). Design of a prediction system for anticipating the consumer's purchase intention of durable goods. International Journal of Microsystems and Iot, 1(6), 418-422. https://doi.org/10.5281/zenodo.10406064

Published online: 27 November 2023.

DOI: https://doi.org/10.5281/zenodo.10406064

International Journal of Microsystems and IoT
Vol. 1, Issue 6, pp. 418- 422; DOI: https://doi.org/10.5281/zenodo.10406064

Check for updates

Design of a prediction system for anticipating the consumer's purchase intention of durable goods

Shaik Qadeer [1], Mohammed Yousuf Khan[2]

[1]Bir Tikendrajit University Manipur, India

ABSTRACT

This study captures a number of shoppers, online searches, and selection state in order to anticipate their purchase intent for durables. Amazon was chosen as an e-commerce platform to collect real-time search and review data from the client. In general, Amazon predicts the customer's purchase intent and promotes the goods in a variety of ways on its website. The suggestion approach was developed with the goal of making the program simple and user-friendly in the e-commerce industry, and research in this subject is still ongoing. Amazon's recommendation algorithms, which have been successful since 2003, employ item-based collaborative filtering. We used the Amazon product database for this study and projection, which contains over 1,500 user reviews for various Amazon items, including the Fire TV Stick, Kindle, and more. The dataset contains basic product characteristics, rating, review text, and more for each product. A powerful e-commerce platform was created for customers by developing a prediction model with attribute level decision support. To build the prediction model, the social perception score of brands and the polarity of feedback are calculated using social network mining and sentiment analysis, respectively. In order to forecast the relevant product attributes for each attribute, a suitable regression analysis and appropriate cases were then constructed for each attribute. In order to use the SVM Algorithm to execute and forecast the model more correctly, we incorporated some additional potent factors, such seasonality and polarity.

KEYWORDS

Support Vector Machine (SVM);
Purchase intent (PI);
Amazon Web Services (AWS);
Machine Learning (ML);
Prediction Model (PM)

1. INTRODUCTION

The trend of online purchasing is remarkably increasing following the emergence of brick and mortar stores. E-retailers generated estimated revenue of 1.9 trillion dollars in 2016 from 1.61 billion customers worldwide, or 7.4% of all retail transactions. The largest online retailer in the world, Amazon, boasts over 310 million active customer accounts and over 136 billion dollars' worth of product that was purchased in 2016 [1]. India, the world's third-buying power equality nation, saw a 271% increase in computerized installment development in the first month of demonetization, but the money down payment fell by around 30–40% at the same time [2]. Furthermore, around 34% of strong products is purchased online by consumers out of the whole market [3]. In light of this, studying how well-founded items are purchased online by customers is a crucial perspective in the online business sector for facilitating seamless online buying. This research has a multidirectional goal. (1) This study provides the product and retailing organization with a comprehensive managerial evaluation. Prediction models will therefore assist consumers in making wise decisions about purchases in a shorter amount of time, and (3) e-retailers may also be able to sustain a platform that provides individualized product recommendations at a reduced cost.

Determining the products that people choose is an interesting task. To identify the clients' choice for every property, a characteristic level forecast model has been suggested in this way.

Fig. 1 Flow Chart of proposed model implementation

The paper's remaining portion is arranged as follows: Section 2 presents the relevant work. Section 3 presents the methods and algorithms. Section 4 presents the results, and Section 5 concludes the research with a review of the work's possibilities for the future.

2. RELATED WORK

This section covers the discussion of the literature review for the proposed research.

The detailed related work in comparison form is shown in Table 1. The table covers the method used to implement the prediction model, their contribution, followed by their advantages and finally disadvantages are listed in tabular form for more clarity.

Table. 1 Literature review of related work

Reference	Method	Advantages	Limitation
[4]	Traditional Word Of Mouth	Social bias is absent from retailer-prompted evaluations, but self-motivated reviews display your social prejudice	This study could only focus on two types of biases, but there certainly are other concerns about online review system
[3]	Propagation tree	Study provides guidelines for luxury brands in virtual shopping perception	In a multi-brand web shop, brands aren't accountable for the store
[2]	Optimum Search	Acquiring digital camera search data reveals evolving search patterns comprehensively.	Augment choice data with search data for better consumer preference identification
[10]	Hybrid data analytic methodology	Optimizing critical data improves transplantation outcome prediction for better heart allocation	Transplantation records hold valuable data on donors, recipients, and procedures
[13]	Error based learning	Explore varied BlockRank measures for informative acquisitions with low computational cost	Researching effective policies for prioritizing acquisitions within the same block.
[1]	Cognitive Approach	Study reveals factors influencing pre-purchase information search for laptops and phones	Study's convenient sampling limits representation of laptop and mobile buyers

3. METHOD AND ALGORITHM

To address the problem of problems mentioned in Table 1, the proposed method concentrate on the following problem statements:

- It is necessary to include a few additional significant criteria in order to improve forecast accuracy
- A general sentiment dictionary based on the Vader rule must be utilized for analysis in order to determine the polarity of the reviews.

The suggested system uses sentiment analysis, social network mining, regression analysis, and big data analysis for data collecting. These analyses can be used alone or in tandem to do more complex analysis.

The massive volume of user-generated information must be cleaned, stored, integrated, and data extracted. A prediction model must then be implemented, and sentiment, social network, and regression analysis must come next.

The suggested prediction model in this research will now include seasonality as a parameter, and based on the findings, the product inventory may be controlled.

The following are the benefits of the suggested work:

- Because the seasonality component was introduced to the project, the prediction model may assist the consumer in making an informed purchase decision in a shorter amount of time.
- An online shop may support a platform that offers clients customized product recommendations.
- This study might provide an abundance of management decisions for e-retailing companies.

The technique utilized is support-vector machines (SVMs, also known as support-vector networks), which are supervised learning models with corresponding learning algorithms that examine the data used for regression analysis and classification in machine learning[14-18]. In comparison to other classifiers like logistic regression and decision trees, SVM delivers extremely high accuracy. It is well-known for its nonlinear input space handling kernel technique. Since our issue involves the categorization of nonlinear spaces, SVM is employed in the model preparation process. Figure 1 shows the flow chart of complete implementation of preparation of proposed prediction model.

The overall flow can be decimated into the following modules:

- Database
- Data Pre-processing
- Polarity Calculation
- Sentiment Analysis
- Data Visualization
- Prediction

Database: Search data from customers and follower IDs of exemplars were gathered from the informs pubs online source. The reports of 74 and 110 exemplars were used in total to determine the brands' SPS (Social Perception Score) for the perceived attributes of luxury and the environment.

Data Pre-processing: Pre-process the data once it has been collected for a database. Pre-processing data simply entails deleting irrelevant information, sanitizing noisy information, and extracting the relevant information. Lastly, the survey data for each item has been cleaned and stored in a separate text file. The code removes rows with missing values, renames some columns, and picks particular columns from the dataset.

Under Feature Engineering it converts the 'reviews_date' column to ‘datetime’ and derives new columns 'reviews_month' and 'season' based on the month of the reviews. Label encoding is applied to the 'brand' and 'categories' columns, replacing categorical values with numerical labels.

Polarity calculation Module: Sentiment analysis has been utilized to determine the polarity ratings of two categories of reviews: product polarity and characteristics level extreme. The various search patterns are connected with the determined extremity scores. For every brand, a social perception score based on the attributes of opulence and eco-friendliness is calculated and then replaced with the brand's name. Every quality that is reported is standardized into a 0–1 range.

Sentiment Analysis: Sentiment analysis is generally realized, and it has determined that the Twitter sentiment of TV shows is extreme in order to predict the TV ratings. The present study employs Vader, a simple guideline-based model for sentiment analysis of information supplied by clients, to arrange the content at three distinct levels: neutral, negative, and positive. Archive-level analysis is being applied at the property level in the present exploration. The record level sentiment analysis has determined the extreme of the whole archive. Alternatively, sentence-level sentiment analysis has been used after the characteristic level sentiment analysis has been applied to the entire archive and reviewed to identify various important interests. The related attribute is now assigned the sentiment score from the comparison sentence. The selection of the attributes was based on the frequency with which the words from the corpus including the entire arrangement of audits pertaining to cameras were distributed. These ascribes are then divided into several perspectives based on similar highlights. Sentiment scores for the 'reviews_text' column are calculated using the AFINN library and stored in a new 'sentiment' column.

Data Visualization: This module plots the sentiment scores using ‘plt.plot(data.sentiment)’ and creates a boxplot of the sentiment scores using ‘plt.boxplot(data.sentiment)’. The code groups the data by 'season' and calculates the mean of different numeric columns. It does the same for 'brand' as well. Bar charts are used to visualize the average values of 'brand,' 'reviews_rating,' and 'sentiment' for different seasons. A similar visualization is performed for 'brand' as well. The code calculates and displays the Pearson correlation matrix for 'brand' and 'season' data. After downloading the ‘NLTK stopwords’, library the code transforms the text input into numerical features using the ‘TfidfVectorizer’ and “TfidfTransformer” library.

Prediction Model: This involves anticipating the purchasing objective of customers based on their quest designs for every attribute. displays the brand's expectation chart as a motivator for extravagance in social perception. The graphic illustrates how nonlinear relapse analysis extraordinarily predicts the actual selected characteristic information. The script prepares the features ('X') and the target variable ('y') for sentiment classification. It also creates a binary 'issed' column based on a threshold sentiment score. The data is split into training and testing sets using the train_test_split function from scikit-learn. Three machine learning models are fitted and evaluated on the data: Linear Regression, Polynomial Regression, and Support Vector Machine (SVM)[14-18]. The accuracy scores of these models on the training data are calculated. A bar chart is created to compare the accuracy scores of the different models.

4. RESULTS

The results of the data processing, feature engineering, and machine learning activities that came before are discussed in this part together with the findings, analyses, and results themselves. Key observations, model performance metrics, and dataset visualizations are probably summarized in this part to lay the groundwork for more analysis and debate. Figure 2 shows the sample of data used for model production

	id	name	asins	brand	categories	keys	manufacturer	reviews.date	revie
0	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
1	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
2	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	
3	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	

Fig. 2 A Sample of data used in the work

After performing data preprocessing and feature engineering, the sample of data now can see in Figure 3.

Out[19]:

	brand	categories	reviews_date	reviews_text	reviews_rating	reviews_month	season	sentiment
0	0	18	2017-01-13 00:00:00+00:00	This product so far has not disappointed. My c...	5.0	1	1	2
1	0	18	2017-01-13 00:00:00+00:00	great for beginner or experienced person. Boug...	5.0	1	1	3
2	0	18	2017-01-13 00:00:00+00:00	Inexpensive tablet for him to use and learn on...	5.0	1	1	2
3	0	18	2017-01-13 00:00:00+00:00	I've had my Fire HD 8 two weeks now and I love...	4.0	1	1	3
4	0	18	2017-01-12 00:00:00+00:00	I bought this for my grand daughter when she c...	5.0	1	1	2
...	...	...	...	...	...	...	...	...
34655	0	8	2012-09-18 00:00:00+00:00	This is not appreciably faster than any other ...	3.0	9	4	1
34656	0	8	2012-11-21 00:00:00+00:00	Amazon should include this charger with the Ki...	1.0	11	4	0
34657	0	8	2012-10-19 00:00:00+00:00	Love my Kindle Fire but I am really disappoint...	1.0	10	4	0
34658	0	8	2012-10-31 00:00:00+00:00	I was surprised to find it did not come with a...	1.0	10	4	0
34659	0	8	2012-12-23 00:00:00+00:00	to spite the fact that i have nothing but good...	1.0	12	1	0

34597 rows × 8 columns

Fig. 3 A Sample of data After preprocessing

It converts the 'reviews_date' column to datetime and derives new columns 'reviews_month' and 'season' based on the month of the reviews. Label encoding is also applied to the 'brand' and 'categories' columns, replacing categorical values with numerical labels. Sentiment scores for the 'reviews_text' column are calculated using the AFINN library and stored in a new 'sentiment' column. The result obtained is shown in Figure 4. Season wise and brand wise analysis is done which is shown in Figure 5 and 6 respectively.

```
plt.plot(data.sentiment)
```

```
[<matplotlib.lines.Line2D at 0x29d34fbda88>]
```

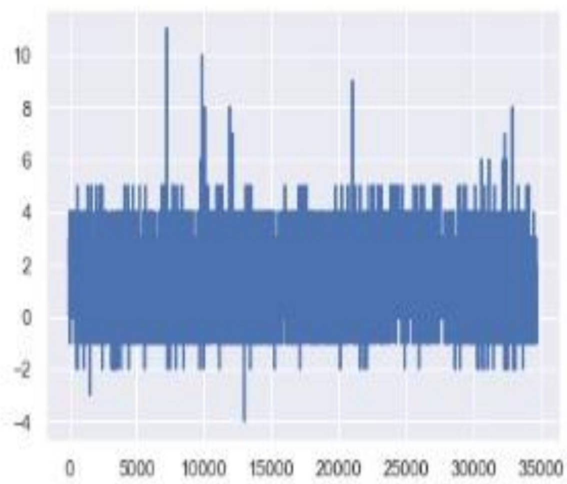

Fig. 4 Sentiment Analysis in Graphical Representation

```
x = season_data.index
ax = plt.subplot(111)
ax.bar(x-0.2, season_data['brand'], width=0.2, color='b', align='center')
ax.bar(x, season_data['reviews_rating'], width=0.2, color='g', align='center')
ax.bar(x+0.2, season_data['sentiment'], width=0.2, color='r', align='center')
plt.show()
```

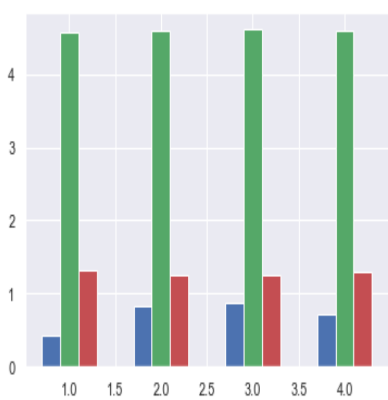

Fig. 5 Season wise Analysis of customer review

```
x = brand_data.index
ax = plt.subplot(111)
ax.bar(x-0.2, brand_data['season'], width=0.2, color='b', align='center')
ax.bar(x, brand_data['reviews_rating'], width=0.2, color='g', align='center')
ax.bar(x+0.2, brand_data['sentiment'], width=0.2, color='r', align='center')
plt.show()
```

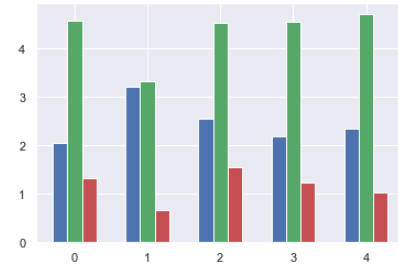

Fig. 6 Brand wise Analysis of customer review

Fig. 7 Accuracy Comparison between proposed SVM and other linear and nonlinear methods.

Finally using the added features, the result obtained by proposed algorithm with the existing method is shown in Figure 7 and Table II. It reveals that, the result obtained by the proposed SVM classification algorithm is better than the existing nonlinear method.

Table. 2 Comparison Of results

Metric	Reading of Passion Pro Tank		
	Linear Method	***Non-Linear Method***	***Proposed SVM***
Accuracy	35%	40%	60%

5. CONCLUSION

Product and online retailers may learn a lot about management from this study. The forecast model's addition of the seasonality element allows clients to make a decision on a purchase in a shorter amount of time. Customers' polarity reviews on Amazon are used in the project to assist users in making decisions on whether or not to buy a product without reading all of the reviews. Moreover, an online store may provide a platform that offers customers customized product recommendations.

Big data analysis may be used in the future in conjunction with relapse analysis, sentiment analysis, interpersonal organization mining, and information assortment, either alone or jointly, to analyse in a more sophisticated way. Specifically, the enormous amount of client-produced content can be extracted, cleaned, stored, and organized using Apache Spark, an open-source group processing framework. The expectation model can then be implemented, and sentiment analysis, unofficial community mining, and relapse analysis can come next. Additionally, certain additional potent elements, like as discounts, rebates, offers, bargains, irregularity, and so forth, may be combined to further enhance the model's efficiency and accuracy.

REFERENCES

1. Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). "Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business", Decision Support Systems, 96, 17–26, https://doi.org/10.1016/j.dss.2017.01.006.
2. Chronicle, D. (2017). The impact of demonetisation on e-commerce. Business, economy.Retrieved from http://www.deccanchronicle.com/business/economy/030317/impactof-demonetisation-on-e-commerce.html.
3. Beuckels, E., &Hudders, L. (2016). An experimental study to investigate the impact of image interactivity on the perception of luxury in an online shopping context. Journal of Retailing and Consumer Services, 33, 135–142.
4. Askalidis, G., Kim, S. J., & Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. Decision Support Systems, 97, 23–30.
5. Baker, A. M., Donthu, N., & Kumar, V. (2016). Investigating how word-of-mouth conversations about brands influence purchase and retransmission intentions. Journal of Marketing Research, 53(2), 225–239.
6. S.A.Z Murad et.al. (2010) High Linearity 5.2GHz CMOS up-conversion Mixer Using Super Derivative Superposition Method, in Proc. TENCON IEEE, 1509-1512.
7. BrightLocal (2016). Local consumer review survey. Retrieved from https://www. brightlocal.com
8. Bronnenberg, B. J., Kim, J. B., &Mela, C. F. (2016). Zooming in on choice: How do consumers search for cameras online? Marketing Science, 35(5), 693–712
9. Chen, J., Teng, L., Yu, Y., & Yu, X. (2016). The effect of online information sources onpurchase intentions between consumers with high and low susceptibility to informationalinfluence. Journal of Business Research, 69(2), 467–475.
10. Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. Marketing Science, 35(3), 343–362
11. Dag, A., Oztekin, A., Yucel, A., Bulur, S., &Megahed, F. M. (2017). Predicting heart transplantation outcomes through data analytics. Decision Support Systems, 94, 42–52.
12. Davvetas, V., & Diamantopoulos, A. (2017). "Regretting your brand-self?" The moderating role of consumer-brand identification on consumer responses to purchase regret. Journal of Business Research, 80, 218–227
13. Deodhar, M., Ghosh, J., Saar-Tsechansky, M., &Keshari, V. (2017). Active learning with multiple localized regression models. INFORMS Journal on Computing, 29(3),503–522
14. S. Q. Mohammad Zakwan and M. Y. khan, "Application of Deep Learning Algorithm in Hydrometry", in international Journal of Hydrology Science and Technology 2024.
15. S. Qadeer, A. G. M, M. Y. Khan and F. Taranum, "Design and Deployment of ML Model for Cardiac Disease Prediction-A review," 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 2023, pp. 1-7, doi: 10.1109/ICIET57285.2023.10220910.
16. M. A. Muqeet, A. B. Mohammad, P. G. Krishna, S. Qadeer and N. Begum, "Automated Oral Cancer Detection using Deep Learning-based Technique," 2022 8th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2022, pp. 294-297, doi: 10.1109/ICSC56524.2022.10009448.
17. M. A. Muqeet, and S. Qadeer "An Empirical Study of CNN-Deep Learning Models for Detection of Covid-19 Using Chest X-Ray Images", Lecture Notes in Networks and Systems, vol 558. Springer, Singapore.
18. Feroza D. Mirajkar; Ruksar Fatima; Shaik A. Qadeer,"Content-based image retrieval using integrated dual deep convolutional neural network", Indonesian Journal of Electrical Engineering and Computer Science, Year 2023 Year 2023,DOI: http://doi.org/10.11591/ijeecs.v31.i1.pp77-87

AUTHORS

Shaik Qadeer received his BE in electronics & instrumentation engineering from Gulbarga University Gulbarga, India 1999 and MTech degree in digital systems and computer electronics from Jawaharlal Nehru Technological University, Hyderabad in 2007 and Ph.D. degree in computer science from Jawaharlal Nehru Technological University Hyderabad, India in 2015. His areas of interest are internet of Things, microcontrollers, PID controller, industrial automation, cyberphysical system, signal processing and machine learning.

Corresponding Author Email: haqbei@gmail.com

Mohammed Yousuf Khan received his BE degree in electronics & instrumentation engineering from Osmania University Hyderabad, India in 1998 and MTech degree in digital systems and computer electronics from Jawaharlal Nehru Technological University, Hyderabad, India in 2004 and Ph.D. degree from Jawaharlal Nehru Technological University Hyderabad, India in 2011. His areas of interest are signal processing in communication, sensor networks and artificial intelligence, computational intelligence etc.

Email: yousuf_lect@yahoo.com