Sentiment analysis on E-commerce reviews and ratings using ML & NLP models to understand consumer behavior

Priyanshi Kathuria
Computer Science Department,
The NorthCap University
Gurugram, India
priyanshikathuria@gmail.com

Parth Sethi
Computer Science Department,
The NorthCap University
Gurugram, India
parth123sethi@gmail.com

Rithwick Negi
Computer Science Department,
The NorthCap University
Gurugram, India
ritwiknegi007@gmail.com

***Abstract*— This research paper covers the sentiment analysis of fashion e-commerce products by comparing their reviews (electronic word-of-mouth), and ratings using the ML model and NLP concepts.**

The scope of the paper dives into the aspect of understanding consumer behavior in a virtual environment. It explores the impact of eWOM and ratings of a product, on customer attitude, and the likelihood of that product being purchased. We have further established a relationship between ratings, reviews, and product recommendations.

The paper contains an exploratory analysis backed with proper reasoning. We have made use of ML classification models like logistic regression, ADA boost, SVM, naïve Bayes, and random forest on customer reviews. We have further used Vader and text blob technologies to perform sentiment analysis.

Keywords*— *ML, NLP, Sentiment analysis, Logistic regression, Ada boost, SVM, Naïve Bayes, random forest, Vader, Text blob, Consumer behavior

I. Introduction

Sentiment analysis has boomed in recent years and has found its usage in most domains of business and real life. Be it movies, management, social media, or politics[17].

This paper is a theoretical piece that contains our conclusions from the application of sentiment analysis of fashion E-commerce products by comparing their reviews (electronic word-of-mouth), and ratings using the ML model and NLP concept.

Sentiment Analytics also known as "Opinion Mining"[18] is the most well-researched concept in most of the major E-commerce brands and social media platforms such as Twitter and Facebook. This is because each of these companies wants to know what emotions consumers emote towards a product/service[19] or issue[1]. The emotion of a consumer, or "eMood"[20] helps companies to understand their target audience better, bring about trends shifts and make calculated and profitable business decisions [7].
A "review" is referred to as "Electronic Word of Mouth (EWOM)" in many places in this paper. It plays an important role in influencing the consumer behavior of users on the Internet. We have approached our dataset in two ways to come to the most appropriate and efficient conclusion about fashion products being purchased online.

1. By applying the concepts of ML and NLP to datasets and drawing our numerical conclusions from the dataset.
2. By syncing our results derived from exploiting the reviews and ratings to establish a brand-new column of product recommendations, to better understand consumer behavior in a virtual environment.

A. Sentiment Analysis

Sentiment analysis is a textual context mining technique that helps companies understand the social sentiment of their brand, product, or service while identifying and extracting subjective information within the source material and monitoring online conversations [3].

Most companies categorize their sentiments into 3 major divisions - good, bad, and neutral (or positive, negative, and neutral). It is imperative for companies to monitor this because a boom in negative sentiment reviews can show the brand in a bad light, which would affect the PR and goodwill of an organization.

Recent advances in deep learning have greatly improved the capabilities of algorithms that analyze text. The creative use of advanced artificial intelligence technology can be a powerful tool for in-depth research. We believe it is important to categorize detailed customer conversations about your brand based on the following lines [11]:

1. An important aspect of the brand's products and services is that the customer is interested in.
2. The user's underlying intent and reaction to these issues.

Combining these basic concepts makes it a very important tool for accurately analyzing the conversations of millions of brands on a human scale.

B. Consumer behavior:

Consumer behavior refers to the psychology, a customer follows whilst buying a product or service. It can include many aspects for example:

1) *Demographics*
2) *Psychographics*
3) *Socio-Cultural factors*

978-1-6654-9604-9/22/$31.00 ©2022 IEEE

4) Age bracket
*5) Product rating ***
*6) Product Review (eWOM) ***

II. Data Description

The 'Women's E-commerce clothing reviews' dataset was used in this study [16]. This dataset consists of all the information related to a review of a particular product that the customer has given. This dataset consists of a mixture of reviews from different ECommerce sites ex- Flipkart, Amazon, etc, ranging from different age groups of people who have given their reviews. The name of the customers who have given a review has been undisclosed. The dataset consists of 23485 rows and 10 columns. The column name is as follows:

1) Age- This column contains the age of the customer who has given a review for a particular product.
2) Title- This column contains the title of the review given.
3) Clothing ID- This column contains the unique ID for that particular product.
4) Division Name - This column contains the category in which the product is segregated.
5) Department Name - This column contains the department name of that particular product in which it is present.
6) Class Name - This column contains the name of the product subcategory i.e pants dresses etc.
7) Review Text - This column contains the review comment that the customer has given to a particular product purchased.
8) Rating - This column contains what rating the customer has given to a particular product purchased.
9) Recommended - This column contains whether the customer has recommended that product to other customers who should buy that particular product or not.
10) Positive Feedback Ct- This column contains a number of other users who have asserted this review.

III. Data Collection and Preprocessing

A. Data Collection

This dataset is a CSV file containing various parameters of customer review analysis.

B. Removing Unforeseen Features

The dataset contained various special characters such as "\n", stopwords, numbers, punctuation, etc in the review text column that needed to be removed for better readability and analysis.

C. Check for Missing Values

"Fig. 1" The dataset columns contained missing values in them. The title and review text had the maximum missing values in them.

```
df.isnull().sum()

Clothing ID                  0
Age                          0
Title                     3810
Review Text                845
Rating                       0
Recommended IND              0
Positive Feedback Count      0
Division Name               14
Department Name             14
Class Name                  14
dtype: int64
```

Fig. 1. List of columns containing the missing value.

We dropped the missing values for the Review Text column for a better understanding and results of our models.

D. Data Exploring

To have a better understanding of our dataset we did an extensive EDA analysis on our dataset to provide us with a better understanding of the dataset because everything in the dataset was in string format specified in the Review Text column. We also used the Word Cloud to analyze the Review Text column along with the Recommended column to distinguish a positive or negative feedback result in the review.

"Fig. 2" represent the positive and negative Word Clouds respectively.

Fig. 2. represents the word cloud

After proper cleaning of the dataset, we moved on to predictive data modeling to draw our conclusions.

IV. Machine learning models on Customer reviews

"Review" in layman's words is a human's sentiment towards a product they have purchased. Sentiments are majorly classified under 3 intensities [13]:

a) Positive
b) Negative
c) Neutral

For our computers to understand the reviews, we apply NLP to simplify our texts.

1) We start with changing text to tokens and converting all of the words to lowercase. Next, we remove punctuation, bad characters, numbers, and stop words. The second step is aimed at normalizing them through the Lemmatization method.

2) Then we extract the stopwords from the NLTK library, and some clothing stopwords and build a function to remove stopwords.

3) We move on to the normalization of words in the corpus by transforming all of the different forms of the word into one. The two methods that exist for this purpose are Stemming and Lemmatization.

4) To apply our ML models, we split our data into test and training sets. In this case, we used only two variables: Review Text and Recommended IND as a predictor and target class variable.

a) "Review Text" is a review from customers and will be used as a predictor variable.
b) "Recommended" is a recommendation from customers, where product 1 is recommended, 0 is not recommended and will be used as the target variable.

5) For bringing out the best results of our models we had tried and tested various hyperparameter tuning methods firstly starting with changing the test size of our <train_test_split>. For our best result 0.2 test size was taken into consideration. Further used some other hyperparameter tuning methods like Grid Search,Random Search and tried our hands on Bayesian Optimization and some regularization techniques.

.

We have used the classification models such as Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, and ADA Boosting. The results are as follows:

Accuracy for Logistic Reg [8]: 88.18%
Accuracy for Naive Bayes [14]: 83.35%
Accuracy for SVM [15]: 80.74%
Accuracy for Random Forest [6][9]: 85.53%
Accuracy for AdaBoost [10]: 85.42%

*Logistic regression model gave us the highest accuracy of all.

V. ML MODELS ON CUSTOMER RATINGS [2][4]

Rating is the numerical grading of an item. On most E-commerce websites, ratings range from 1 to 5; where 1 means bad quality/least satisfaction and 5 means good quality/complete satisfaction.

The rating range bracket is as follows:

a. 4-5 = Positive
b. 3 = Neutral
c. 1-2 = Negative

We have drawn a strong relationship between ratings and product recommendations. Where 1 means that the customer has further recommended using that product to another customer, and 0 means that the product has not been recommended to be used.

For our computers to understand the reviews, we apply NLP to simplify our texts.

1) We have created a table with a comparison between ratings, reviews, and recommendations.

2) We have converted our ratings into 3 classes: Positive, neutral, and negative.

3) We start with changing text to tokens and converting all of the words to lowercase.

4) Next, we remove punctuation, bad characters, numbers, and stop words. The second step is aimed at normalizing them through the Lemmatization method.

5) We move on to the normalization of words in the corpus by transforming all of the different forms of the word into one. The two methods that exist for this purpose are Stemming and Lemmatization.

6) To apply our ML models, we split our data into test and training sets. In this case, we used only two variables: Review **Text** and **Class** as predictor and target class variables.

a) "Review Text" is a review from customers and will be used as a predictor variable.
b) "Class" is the rating bracket, where:
   - Positive = 4-5
   - Neutral = 3
   - Negative = 1-2

7) For bringing out the best results of our models we had tried and tested various hyperparameter tuning methods firstly starting with changing the test size of our <train_test_split>. For our best result 0.2 test size was taken into consideration. Further used some other hyperparameter tuning methods like Grid Search,Random Search and tried our hands on Bayesian Optimization and some regularization techniques.

We have used the classification models such as Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, and ADA Boosting. The results are as follows:

Accuracy for Logistic Reg [8]: 80.68%
Accuracy for Naive Bayes [14]: 75.99%
Accuracy for SVM [15]: 80.30%
Accuracy for Random Forest [6][9]: 77.98%
Accuracy for AdaBoost [10]: 78.94 %

*Logistic regression model gave us the highest accuracy of all.

VI. VADER AND TEXTBLOB FOR SENTIMENT ANALYSIS

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed by a user [12].

As TextBlob is a Lexicon-based sentiment analyzer It has some predefined rules or we can say word and weight dictionary, where it has some scores that help to calculate a sentence's polarity.

1) Firstly for VADER analysis, we created a minimized data frame for our analysis i.e consisting of rating, review, and recommendation. Subsequently, we categorized the reviews given corresponding to class results i.e. positive, negative, and neutral. If the rating given is more than 4 then positive, 3 means neutral, and less than 3 is negative.

2) Then we calculated the lexicon scores[19] for the reviews present in our dataset [5]. Positive, negative, and neutral results represent the proportion of the text that falls into these categories. This means a review that our opinion was rated as 79% neutral, 18% positive, and 0.3% negative. All results should add up to 1. The Compound score is a metric that calculates the sum of all lexicon scores that have been normalized between -1 (most extreme negative) and +1 (most extreme positive). The compound is 0.94, which means a very high positive sentiment. Each review has a negative, neutral, positive, and complex result. The complex result is a comprehensive assessment of the first three points and this score is in the range of -1 to 1. Based on these results, we can determine the sentiment. We set the score threshold ourselves and here we can set it to ± 0.5. If the compound is bigger than 0.5 the review is positive, from 0 to 0.5 is neutral, and below 0 is negative

3) Based on the compound score generated by the process the categorized the sentiment of the review. If the compound score is greater than 0.5, it's a positive sentiment, if in between 0 to 0.5 it's a neutral sentiment and if less than 0 it's a negative sentiment.

4) We received an accuracy of 76% for this. The Vader assigned more positive reviews than the original rating, making the study progress.

5) Secondly, for TEXT BLOB analysis we created a similarly minimized data frame similar to what we did for VADER along with categorizing the class for different reviews given.

6) Then we calculated the polarity and subjectivity of a sentence.

Eg- The polarity of the sentence is 0.63, indicating that the sentiment is positive. The subjectivity of the text is 0.93 in our example. A value closer to 1 indicates that the sentence is mostly a public opinion and not a factual piece of information and vice versa.|

7) Subsequently, we categorized the polarity scores to sentiment results if the polarity score is greater than 0 is a positive sentiment, less than 0 its a negative sentiment else neutral.

8) We received an accuracy score of 77% for TEXT BLOB. This method assigned even more positive reviews than VADER and normal technique making the study even more progressive.

VII. Comparison of models

"TABLE 1" The following table is a comparative study between all the models and elements used in the research paper. You can evaluate the highest accuracy:

TABLE 1.
Model accuracy comparison.

Customer reviews	**Vader & TextBlob**	**Customer rating**
Highest accuracy (Logistic Reg): **88.18%**	Vadar = **76%** TextBlob = **77%**	Highest accuracy (Logistic Reg): **80.68%**

VIII. Conclusion

This research paper is a comparative study on consumer behavior analysis where we used different parameters to understand how a consumer decides whether they like a product or not. We focused on product reviews, ratings, and recommendations to draw conclusions. We used ML techniques like logistic regression, random forest, Naive Bayes, SVM, Random Forest, and AdaBoost to draw our accuracies. We also used Vader and Text Blob technologies for sentiment analysis. The sentiment analysis that we have done is the outcome of how positive reviews were given by different users or customers while purchasing any product. As mentioned, we categorized the reviews along with ratings and also with the recommendation of the product to analyze for different ML models, Text Blob and Vader. We got to analyze that Logistic Regression had the best results out of all ML models giving the highest accuracy for our model. Subsequently, we also analyzed how the sentiment analysis of reviews can help to know how the product has shaped its market league. It also helps in understanding the product listed on the e-commerce website, and how well it attracts the audience with its rating and reviews which eventually helps the other customers or users identify the right choice for purchasing. It also helps to give feedback to the seller if their product is not doing well, then what changes they need to make to improve their results in the market.

IX. References

[1] Gao, Yuan, ed. *Web systems design and online consumer behavior*. IGI Global, 2005

[2] Raza, Hassan, et al. "Scientific text sentiment analysis using machine learning techniques." *International Journal of Advanced Computer Science and Applications* 10.12 (2019): 157-165.

[3] Nawaz, Zubair, et al. "Role of Artificial Neural Networks Techniques in Development of Market Intelligence: A Study of Sentiment Analysis of eWOM of a Women's Clothing Company." *Journal of Theoretical and Applied Electronic Commerce Research* 16.5 (2021): 1862-1873

[4] Xie, Shuangyin. *Sentiment Analysis using machine learning algorithms: online women clothing reviews*. Diss. Dublin, National College of Ireland, 2019.

[5] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5.4 (2014): 1093-1113.

[6] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

[7] Khwaja, Muddasar Ghani, and Umer Zaman. "Configuring the evolving role of ewom on the consumers information adoption." *Journal of Open Innovation: Technology, Market, and Complexity* 6.4 (2020): 125.

[8] Peng, Chao-Ying Joanne, Kuk Lida Lee, and Gary M. Ingersoll. "An introduction to logistic regression analysis and reporting." *The journal of educational research* 96.1 (2002): 3-14.
[9] Biau, Gérard. "Analysis of a random forest model." *The Journal of Machine Learning Research* 13.1 (2012): 1063-1095.
[10] Chengsheng, Tu, Liu Huacheng, and Xu Bing. "AdaBoost typical Algorithm and its application research." *MATEC Web of Conferences*. Vol. 139. EDP Sciences, 2017.
[11] Poria, Soujanya, et al. "Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research." *IEEE Transactions on Affective Computing* (2020).
[12] Hutto, Clayton, and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the international AAAI conference on web and social media*. Vol. 8. No. 1. 2014.
[13] Mitra, Ayushi. "Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset)." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2.03 (2020): 145-152
[14] Kaviani, Pouria, and Sunita Dhotre. "Short survey on naive bayes algorithm." *International Journal of Advance Engineering and Research Development* 4.11 (2017): 607-611.
[15] Evgeniou, Theodoros, and Massimiliano Pontil. "Support vector machines: Theory and applications." *Advanced Course on Artificial Intelligence*. Springer, Berlin, Heidelberg, 1999.
[16] Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network, Abien Fred Agarap, Women's E-commerce clothing reviews, https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews
[17] Priyavrat Chauhan, Nonita Sharma, and Geeta Sikka. "The emergence pf social media data and sentiment analysis in election prediction." Journal of Ambient Intelligence and Humanized Computing, 2021.
[18] Marouane Birjali, Mohammed Kasri, and Abderrahim Beni-Hssane. "A comprehensive survey on sentiment analysis: Approaches, Challenges and Trends." LAROSERI Laboratory, Computer Science Department, University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco, 2021.
[19] M.B. Nasreen Taj and G.S.Girisha. "Insights of strength and weakness of evolving methodologies of sentiment analysis." Dayananda Sagar University, Kudulugate, Karnataka, India and Department of Information Science and Engineering, GMIT, Davanagere, Karnataka, India, 2021.
[20] A.H. Alamoodi, B.B. Zaidan, A.A. Zaidan, O.S. Albahri, K.I. Mohammed, R.Q. Malik, E.M. Almahdi, M.A. Chyad, Z. Tareq, A.S. Albahri, Hamsa Hameed, Musaab Alaa. "Sentiment analysis and its application in fighting COVID-19 and infectious diseases: A systematic review." Department of Computing, Sultan Idris University of Education (UPSI), Tanjong Malim, Malaysia, 2020.