

Consumer Opinion Analysis:A Data-Based Examination

C H Anil kumar, M M S pavan, K.Gnana Deep, R.Manjunath, Amulyashree Sridhar

Department of Computer Science and Engineering

Amrita School Of Computing, Bengaluru

Amrita Vishwa Vidyapeetham, India,

bl.en.u4cse22213@bl.students.amrita.edu, bl.en.u4cse22235@bl.students.amrita.edu, bl.en.u4cse22228@bl.students.amrita.edu, bl.en.u4cse22277@bl.students.amrita.edu, s_amulyashree@blr.amrita.edu

Abstract—Consumer reviews and opinions play a very important role in making or breaking the success of the different e-commerce platforms. The study deals with extracting customer reviews from leading e-commerce websites including Flipkart, Amazon, Fresh India Organics, and BigBasket in order to mine useful insights into customer sentiments. For analyzing these customer reviews, a whole pipeline data analysis framework has been implemented, which includes emotional analysis, NER, N-gram and cluster analysis. Share of Positive Sentiments Towards Products Attributes The results have a high presence of positive sentiments, especially those related to product attributes

Keywords—E-commerce, Consumer Feedback, Sentiment Analysis, Emotional Analysis, NER, N-gram Analysis, Cluster Analysis

I. INTRODUCTION

Fruits form the backbone of human diet and nutrition in recent decades because they contain vitamins, minerals, and fibers as their quintessential constituents. The global fruit sector is steered by increasing awareness about health benefits and sustainable living. The fruits figure in the diet as well as agriculture of India, which is considered one of the largest producers of mangoes, bananas, and guavas. In addition, the global fruit trade has also shown a steady rise, with tropical fruits like avocados and dragon fruits gaining great momentum in popularity across a number of markets. Fruits are nutritionally very important since they acting as antioxidants to prevent chronic diseases and improve one's general health; therefore, they cannot be dispensed with in diets. Their consumption is actually part of the emerging trends in health, which supports the physiological and economic health of nations worldwide.

The author talks about the impact of eWOMs on the decision-making of consumers has been one of the most popular themes studied. For example, in one study, a few survival strategies for family-run homestays were identified; customer reviews also create perceptions and help improve services through text mining techniques [1]. Another discussed how consumer behavior is evolving and trending, and that eWOM plays an important role in the purchasing decision-making process, derived from insights of major consumer behavior journals [2]. The online reviews and their impacts on tourist satisfaction in wine tourism using the lens of experiential feedback were studied, which showed critical observations that

can help understand customer intention in the e-travel sector [3]. In a related study, customer reviews in the context of e-travel services have been sought to shed light on service process innovations, considering how active product reviews can be in the improvement of service.

Further, one study looked at the perceived information quality obtained from online reviews and focused on the credibility of the reviews in shaping consumer preferences [5]. Viral marketing, for being an effective advertising tool, has also been studied wherein eWOM has come across as an important contributor in brand building and influencing Gen-Y purchase intentions for India [6]. These pieces of studies collectively establish the role of electronic word-of-mouths in deciphering customer preferences, driving marketing strategies, and enhancing customer satisfaction for any industry..

The inferences from these studies point out that the current study focuses on analyzing reviews made by consumers with respect to fruits' popularity, progressing with an understanding of consumer preferences and emotional responses associated with different types of fruits. In analyzing knowledge from e-WoMs, data from various online review platforms were processed using a framework for data analysis. The contributions of this research work are listed below:

- Collected review data of different fruits from online review platforms; preprocessed this data by tokenization, removing stop-words, and text normalization (lowercasing).
- Visualized data collected with the help of Word Cloud, which enabled the identification of the most frequently occurring words in reviews.
- Performed emotional analysis to classify reviews into eight categories of emotions and three classes of sentiment: positive, neutral, and negative.
- Categorized the entities extracted from the text corpus using Named Entity Recognition (NER) techniques.
- Interpreted the sequence of words using *N*-gram analysis to identify common phrases and associations in reviews.
- Mapped the relationship of words to sentiments using hierarchical clustering on consumer feedback to find patterns.

II. RELATED WORKS

This section highlights prominent studies on sentiment analysis in e-commerce and customer feedback. Initial study reviewed methods for sentiment analysis of e-commerce customer reviews using machine learning techniques. It emphasized the influence of customer reviews on buying behavior and classified reviews into positive, neutral, and negative sentiments using methods like Naive Bayes and Support Vector Machines (SVM) [7]. Another study evaluated how sentiment analysis can enhance product recommendations by analyzing customer reviews, achieving high accuracy rates (80-90%) with Logistic Regression performing best [8].

Another study focused on applying sentiment analysis to improve product recommendations in e-commerce, highlighting the significance of online reviews in influencing purchasing decisions [9]. Another study explored sentiment analysis techniques for recommendation systems, evaluating methods like SVM and deep learning models, achieving high accuracy in predicting customer preferences [10].

Another study analyzed customer feedback data from Amazon product reviews, emphasizing the importance of understanding consumer sentiments for improving recommendations [11]. A study titled "Sentiment Analysis using VADER" analyzed British Airways customer reviews, revealing a predominance of positive sentiments [12]. Another study investigated the effectiveness of SVM and CNN in analyzing customer sentiment, achieving high accuracy through dataset cleaning and preprocessing [13].

Another study presented a methodology for analyzing restaurant reviews using Natural Language Processing (NLP) techniques, emphasizing customer satisfaction [14]. Another study discussed the significance of analyzing customer sentiment through online reviews, proposing text mining methods and sentiment scores as alternative metrics [15]. Another study revealed that SVM outperformed CNN in sentiment analysis accuracy [16].

Another study proposed a model for analyzing customer satisfaction through text reviews, utilizing text mining techniques to extract key factors [17]. Another study examined factors affecting customer satisfaction in delivery applications, identifying accessibility and utility as significant elements [18]. Another study discussed the role of computer vision in inspecting fruits and vegetables, emphasizing visual appearance in customer preferences [19].

Another study explored aspect sentiment analysis combined with triplet extraction to provide insights into customer preferences in e-commerce [20]. Another study identified critical factors affecting customer satisfaction in online agricultural purchases, focusing on product attributes and services [21]. Lastly, a study presented a novel approach to extracting customer requirements from online reviews using multi-aspected sentiment analysis and the KANO model [22].

III. METHODOLOGY

The following section describes the methodology of the proposed work as visualized in Fig.1.

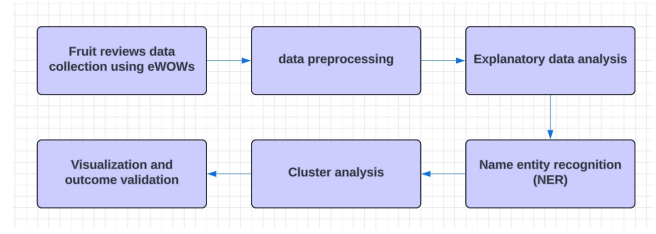


Fig. 1. Analytical flowchart of the present study

A. Data Collection

1) *Web Scrapping*: The reviews of several products from e-commerce portals are collected; these include Amazon, Flipkart, FreshIndiaOrganics, and BigBasket. All the reviews collected are directly from the above-mentioned portals, while only verified feedback from registered customers is considered for maximum reliability. Hence, all the collected reviews were organized and saved in a CSV format, making it easier to handle and analyze. Three key features were recorded for each review: data source (the source of the review, namely, which platform it was taken from), review text (the content of the feedback), and label- Positive, Negative, or Neutral.. The description of data collected is represented in Table 1.

TABLE I
DESCRIPTION OF DATA COLLECTED

Type	Data Instances	Data Source	Missing Values?	Label	Duration of Data Collection
Textual	750	Amazon, Flipkart, FreshIndiaOrganics, and BigBasket	No	Yes	1st Aug 2024 - 15th Aug 2024

B. Data Preprocessing

The data collected is further pre-processed to eliminate irrelevant content, which is not essential for further analysis. In this perspective, the data was subjected to preprocessing techniques including, elimination of English stopwords, elimination of digits, and special symbols, followed by elimination of punctuation symbols. Also, data is converted to lower case, followed by tokenization and stemming to derive tokens and stemmed words. The pre-processed data is termed as 'text corpus' for further analysis

TABLE II
SAMPLE OF EWOMS DATA COLLECTED

Sl.No	Comment	Label	Data Source
1	Very good in taste and awesome	Positive	Amazon
2	Didn't meet the expectations as shown in photos	Negative	Flipkart
3	Bitter couldn't eat and also very bad in taste	Negative	FreshIndiaorg
4	Not worst but average taste	Neutral	Bigbasket
5	cheap and best	positive	Flipkart

- 1) Converting to Lowercase: Preprocessing the dataset converts all of a document to lowercase. This means that words such as "Great" and "great" are the same.
- 2) Tokenization: It is a breaking down of a comment into individual words or subunits known as tokens. Within this analysis, these tokens are used for either further analysis or feature extraction of the sentiment analysis model.
- 3) Stop word Elimination: Very common stop words are words like "is," "the," "a," "an," and "in," referring to the common meaning that would add rarely or insignificantly in giving information, and because of this, are usually cut down to reduce noise. It reduces the dimensionality of a dataset and thus makes the model function better by limiting consideration of useful words.

C. Emotional Analysis

The preprocessed analyzed data, the level of satisfaction or dissatisfaction about different fruits can be judged. The positive emotions express satisfaction about the fruits, while through neutral emotions, there is no expression of any specific opinion about fruits. On the other side, negative emotions arise because of dissatisfaction. The research uses supervised classification where three emotion levels include 1 = Positive, 0 = Neutral, and -1 = Negative. The emotionally satisfied reviewers will like to recommend these fruits; hence, exploring the emotional tones in the reviews becomes important. It performs word emotional analysis with the help of the R module 'syuzhet' and categorizes emotions into eight distinct types: trust, joy, anticipation, surprise, fear, anger, disgust, and sadness. The positive feelings mean that one enjoys or trusts the fruit's quality, the neutral feelings show no concern for fruit, and the negative feelings reflect aversion to it. The insight drawn from here gives the customer preference and emotional response to the fruit, which will be helpful in product offering or market strategy.

D. Exploratory analysis:

To identify relevant terms within the text corpus, the term frequency (TF) and Bag-of-Words (BoW) were estimated. Here, habitually occurring words are identified from corpus. TF is mathematically represented in Eq.1:

$$TF(x_a, y_b) = \frac{\text{Frequency of } x_a \text{ in } y_b}{\text{Total number of terms in } y_b} \quad (1)$$

Here, x_a is the term that is present in the document y_b .

Additionally, the BoW approach is utilized to identify relevant terms within the corpus. Here, the terms are represented in a "bag," thereby deriving prescribed length vector representation from eWOMs. Thereby, text data is transformed into vector-encapsulating digits.

E. Name Entity Recognition(NER)

This approach is commonly used in processing text by ascertaining and categorizing entities into pre-specified entities based on categories. In this context, NER is performed on the text corpus to recognize entities.

F. N-gram Analysis

N-gram analysis is an approach in text analysis used to interpret sequences of n words within a corpus. This technique partitions the text into bi-grams or 2-word sequences, followed by tri-grams or 3-word sequences, thus identifying inherent word patterns and association amongst multiple words.

G. Text Clustering

To identify interconnections between terms within the corpus, the current study employed a hierarchical clustering approach. In this perspective, the Euclidean distance matrix is computed to ascertain interconnections across the terms. Here, a complete linkage technique is applied to derive the hierarchical clusters [28].

IV. RESULTS AND DISCUSSIONS

A. Outcomes from Emotional Analysis

Outcomes from the emotional analysis are visualized in Fig.2. As visualized from the figure, the factor of trust was

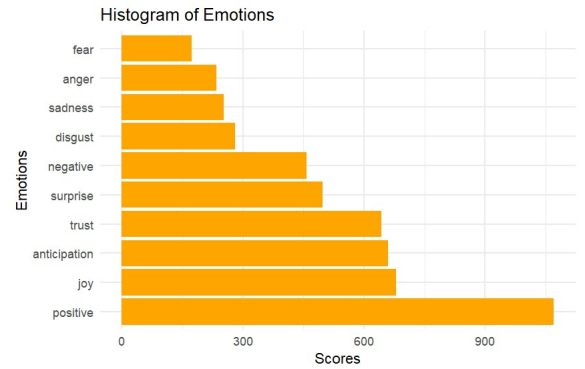


Fig. 2. Emotion Analysis

the significant emotional insight from learners with highest score. It was followed by joy and anticipation. On the contrary, disgust was the least expressed emotion. These outcomes reveal that learners were well pleased with the content from the influencer and anticipated towards further content from the influencer.

B. Outcomes from Exploratory Analysis

Online reviews from the influencer were analyzed to identify frequently occurring terms. These terms are visualized in Fig.3. It is observed from the figure that "video", "math", "beautiful", "like", and "amazing" were significant words from the corpus. These words represent positive impressions from learners towards the influencer. Also, Word cloud was visualized from reviews towards the influencer as reflected in Fig.4. As observed, the words "Thank", "video", "math", "understand", and "beautiful" were significantly expressed, indicating positive impressions towards the influencer.

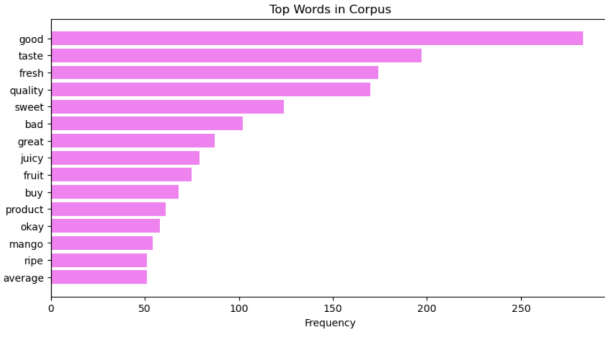


Fig. 3. Top 10 words identified from the corpus



Fig. 4. Word cloud identified from data

C. Name Entity Recognition

Outcomes from NER analysis are reflected in Table 3. As observed from the table, entities including "Like", "Thank", "Amazing", "Math" and "Video" were significantly observed from eWOMs on the influencer. These outcomes also reflected pragmatic opinions towards the influencer.

TABLE III
CATEGORIZED ENTITIES FROM NER ANALYSIS

Entity Name	Frequency	Category of Entity
Good	270	Adjective
Taste	210	Noun
Fresh	160	Adjective
Quality	150	Noun
Sweet	120	Adjective
Bad	110	Adjective
Great	100	Adjective
Juicy	90	Adjective
Fruit	80	Noun
Buy	70	Verb
Product	60	Noun
Okay	50	Adjective
Mango	50	Noun
Ripe	50	Adjective
Average	40	Adjective

D. Outcomes from N-gram analysis

Experiments were conducted to identify associations between terms using N-gram analysis. Primarily, when $n=1$, no prominent associations were identified. Furthermore, when $n=2$ and $n=3$, significant relationships were detected as visualized in Fig.5 and Fig.6 respectively. Both the figures identified top 20 associations with "linear algebra" and "course", being

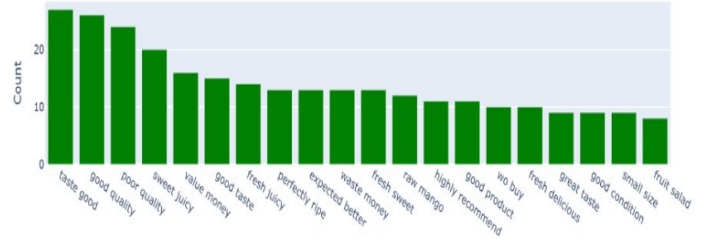


Fig. 5. Outcomes from Bigram analysis

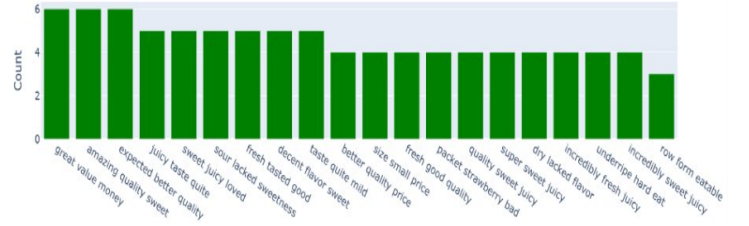


Fig. 6. Outcomes from Trigram analysis

reflected due to the popularity of linear algebra courses by the influencer.

E. Outcomes from cluster analysis

Hierarchical cluster analysis was performed to identify knowledgeable insights on consumer satisfaction. The dendrogram derived is visualized in Fig.7. It indicated an inverse relationship with distance and terms. The terms "thank", "love", "amazing", "good", "understand", "great", and "thanks" from multiple clusters revealed constructive opinions towards the influencer.

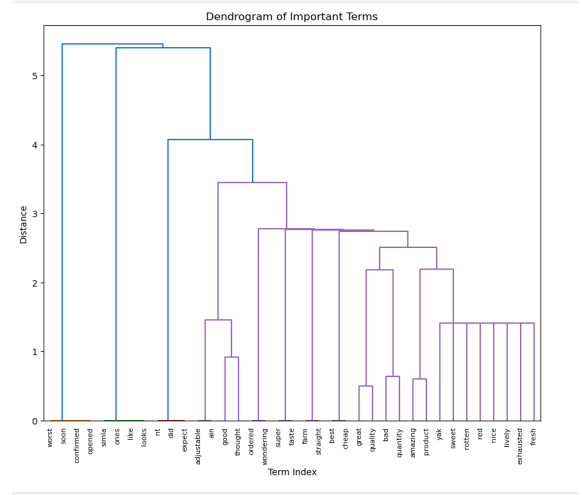


Fig. 7. Dendrogram identified from cluster analysis

V. CONCLUSION

This study analyzed the consumer opinions regarding several fruits and performed structured analysis based on emotions, NER, N-gram analysis, and hierarchical clustering. The

results showed that the consumers have positive attitudes toward the products because majority of frequently occurring terms like "fresh," "delicious," "sweet," and "healthy" were found in the text corpus. Such positive perceptions could be due to nutritional and freshness qualities and favorable taste of the fruits that were under review. Negative attitudes were few in number and largely an issue of worries about packaging or freshness. Cluster analysis further confirmed this, as it focused on clusters that boasted satisfaction and appreciation by consumers.

However, the current study had limitations. As it was based on specific fruits, the sample was homogeneous and thus could not be generalized to other categories. It was also collected over a fixed time frame, hence longitudinal analysis could not be provided. There is scope for a much greater dataset with more diversified types of fruits and time periods that may actually help in deeper insights regarding consumer preferences for improving marketing strategies and customer satisfaction.

REFERENCES

- [1] J. Krishnan, B. Bhattacharjee, M. Pratap, J. K. Yadav, and M. Maiti, "Survival strategies for family-run homestays: analyzing user reviews through text mining," *Data Science and Management*, vol. 7, no. 3, pp. 228–237, Sep. 2024, doi: 10.1016/j.dsm.2024.03.003.
- [2] W. M. Lim, S. Kumar, N. Pandey, D. Verma, and D. Kumar, "Evolution and trends in consumer behaviour: Insights from Journal of Consumer Behaviour," *Journal of Consumer Behaviour*, vol. 22, no. 1, pp. 217–232, Jan. 2023, doi: 10.1002/cb.2118.
- [3] S. Gunasekar, P. Das, S. K. Dixit, S. Mandal, and S. R. Mehta, "Wine-experiance and tourist satisfaction: through the lens of online reviews," *Journal of Foodservice Business Research*, vol. 25, no. 6, pp. 684–701, Nov. 2022, doi: 10.1080/15378020.2021.2006039.
- [4] G. Rejikumar, V. Nimisha, and S. V. Mohan, "An examination of customer reviews for service process innovations: A study with respect to E-travel services," in *2019 11th International Conference on Advanced Computing (ICoAC)*, IEEE, Dec. 2019, pp. 152–157, doi: 10.1109/ICoAC48765.2019.246832.
- [5] J. Gobinath and D. Gupta, "Online reviews: Determining the perceived quality of information," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, Sep. 2016, pp. 412–416, doi: 10.1109/ICACCI.2016.7732080.
- [6] U. Rawat and R. Prasad, "Is Viral Marketing an effective and reliable method of advertising and branding? A perspective of Gen- Y of India," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, Aug. 2015, pp. 1839–1842, doi: 10.1109/ICACCI.2015.7275885.
- [7] Seyfioglu, Mehmet Saygin, and Mustafa Umut Demirezen, "A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data," *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2017.
- [8] Im, Eun Tack, et al. "A study on the extraction of customer satisfaction factors based on the customer satisfaction model using text review and preview," *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, IEEE, 2021.
- [9] Lee, Seungwook, Weonsun Choi, and Jaeyoung So, "A Study on the Factors Affecting Customer Satisfaction in Delivery Applications: Focusing on Sentiment Analysis of Review Data," *2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD)*, IEEE, 2022.
- [10] Prabhu, Akshatha, et al. "Applications of computer vision for defect detection in fruits: A review," *2021 International Conference on Intelligent Technologies (CONIT)*, IEEE, 2021.
- [11] Princess, P. Joyce Beryl, "Aspect Sentiment Analysis with Triplet Extraction for E-Commerce," *2024 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, 2024.
- [12] Liu, Yimin, "Customer review elements of agricultural products e-commerce and their influence on satisfaction based on big data crawlers," *2021 2nd International Conference on Big Data Economy and Information Management (BDEIM)*, IEEE, 2021.
- [13] Jiang, Kaicheng, and Yanting Li, "Mining customer requirement from online reviews based on multi-aspected sentiment analysis and Kano model," *2020 16th Dahe Fortune China Forum and Chinese High-educational Management Annual Academic Conference (DFHMC)*, IEEE, 2020.
- [14] Bharti, P. K., et al. "Review of Efficient Methods for Sentiment Analysis of E-Commerce Customer Review using Machine Learning Techniques," *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, 2023.
- [15] Panduro-Ramirez, Jeidy, "Sentiment Analysis in Customer Reviews for Product Recommendation in E-commerce Using Machine Learning," *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, IEEE, 2024.
- [16] Kanakamedala, Venkata Chaitanya, Sonal Hukampal Singh, and Rishitha Talasani, "Sentiment analysis of online customer reviews for handicraft product using machine learning: A case of Flipkart," *2023 International Conference for Advancement in Technology (ICONAT)*, IEEE, 2023.
- [17] Laksono, Rachmawan Adi, et al. "Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes," *2019 12th international conference on information communication technology and system (ICTS)*, IEEE, 2019.
- [18] Pandey, Prashant, and Nitasha Soni, "Sentiment analysis on customer feedback data: Amazon product reviews," *2019 International conference on machine learning, big data, cloud and parallel computing (COMIT-Con)*, IEEE, 2019.
- [19] Annamalai, R., et al. "Sentiment Analysis using VADER: Unveiling Customer Sentiment and Predicting Buying Behavior in the Airline Industry," *2024 IEEE International Conference for Women in Innovation, Technology Entrepreneurship (ICWITE)*, IEEE, 2024.
- [20] Uma, R., Priyanka Jawahar, and Bugatha Venkata Rishitha, "Support Vector Machine and Convolutional Neural Network Approach to Customer Review Sentiment Analysis," *2022 1st International Conference on Computational Science and Technology (ICCST)*, IEEE, 2022.
- [21] Ara, Jinat, et al. "Understanding customer sentiment: Lexical analysis of restaurant reviews," *2020 IEEE Region 10 Symposium (TENSYP)*, IEEE, 2020.
- [22] Kim, Rae Yule, "Using online reviews for customer sentiment analysis," *IEEE Engineering Management Review* 49.4 (2021): 162-168.