Health Benefits and Consumer Sentiment: Analyzing Edible Seed Reviews

Ojas, G. Meher Pranav Narayan, Saurabh Kiran Mahuli, B. Uma Maheswari, Amulyashree Sridhar
Department of Computer Science and Engineering
Amrita School of Computing, Bangalore
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22281@bl.students.amrita.edu, bl.en.u4cse22223@bl.students.amrita.edu, bl.en.u4cse2253@bl.students.amrita.edu,
b_uma@blr.amrita.edu, s_amulyashree@blr.amrita.edu

***Abstract*—In this research, the consumer perception of edible seeds is explored by using online consumer reviews obtained from online platforms including Amazon, BigBasket, Flipkart, and YouTube. The first step of the analysis is Data Preprocessing, where the text data is tokenized and non-alphanumeric characters are removed from the text data while the Named Entity Recognition (NER) algorithm is applied to identify important features such as brands and products as mentioned in the reviews. Regarding the classification, we use Agglomerative and hierarchical clustering to categorize the comments' sentiment. Such an approach allows us to find associations between the discussions of sentiment analysis with the literature on chia, sunflower, and sesame seed's health impact. The present study intends to look at consumer attitudes and behavior toward edible seeds and help brands better understand how to engage customers by modifying their marketing strategies.**

***Index Terms*—Text Clustering, electronic Word-of-Mouth, Sentiment Analysis, Online Shopping, Named Entity Recognition**

I. Introduction

The edible seeds market has also been on the rise in recent years due to the growing concern for healthy food by consumers and the growing shift by humans from animal-based food products. Some of the common seeds and nuts that are eaten include chia, flaxseed, pumpkin seeds, and sunflower seeds, they are now known to contain high amounts of omega-3 fatty acids, protein, fiber, and other vitamins and minerals. The increasing awareness of these benefits has also made consumers use these products for cooking and baking besides the health benefits. Promoting edible seeds requires creating recognition of such products therefore, health, nutrition, and wellness Internet comments are ideal for this practice because they provide reviews about these edible seeds and their experiences. This form of marketing focuses on the use of the message in the form of electronic word-of-mouth (eWOM) in which referrals from influential sources greatly shift consumer decision-making. Cite sources that show that customers are more likely to listen to recommendations on products sold online and their percentage is at 93% [1].

The pressure created by eWOM communication is more significant and more influential in high-involvement product categories such as edible seeds, where the consumers are proactively looking for detailed information about the product they are planning to buy. Consumer reviews and recommendations by opinion leaders can improve consumer perception and increase purchase intention. On the other hand, negative feedback can repel possible buyers and change their attitude toward the quality of the product [4]. Hence, knowledge of the unfolding trends in the edible seeds market specifically regarding Customer ratings is potentially valuable for those brands with an eye on the said market trend. This paper aims to establish structural relationships between consumer sentiment and buying tendencies of edible seeds by examining reviews from prominent e-marketplaces and using sentiment analysis. The study's results will advance the literature on influencing marketing by consumers and the influence of consumer decision-making patterns.

II. Related Works

A. Electronic word of mouth(eWOM)

Electronic Word of Mouth (eWOM) plays a pivotal role in consumer decision-making, particularly on social media platforms. Unlike traditional word-of-mouth communication, eWOM allows users to instantly share their thoughts, experiences, and reviews with a vast, global audience. Platforms like Facebook, Instagram, and Twitter enable consumers to influence others' purchasing decisions by sharing their personal experiences with a wider network[6]. Research suggests that consumers perceive eWOM as more trustworthy compared to traditional advertising, as it often comes from impartial third-party sources. This increased credibility fosters greater trust and engagement, making eWOM a powerful tool in shaping consumer opinions and purchase behavior[8]. The interactive features of social media further amplify the impact of eWOM, as users can like, comment, or share posts, helping these messages reach an even broader audience[3]. Moreover, studies show that eWOM's influence is stronger from personal connections such as friends or family, as consumers are more inclined to trust recommendations from within their social circles[9]. Given this influence, businesses increasingly prioritize strategies that encourage positive eWOM, as it enhances brand reputation and drives consumer engagement in the digital marketplace. Perceptions of eWOMs have been reviewed from various domains to unearth concealed trends, especially regarding consumers' behavior, service quality, and product satisfaction. For instance, one research focused on analyzing the customers' feedback regarding the family mart

homestays in Kerala by incorporating sentiment analysis, entity analysis and text mining tools for extracting important topics like cleanliness, hospitality, etc, for enjoying a very high accuracy in the logistic regression predictive models[11]. Similarly, another study also performed a bibliometric analysis of consumer behavior research, highlighting the increasing factor of external influence like COVID-19 in consumers' behavior and their relationship with brands[12].

Other scholars, including a study on tourist satisfaction at Indian wineries[13] and profitability of e-travel services[14], have shown how eWOM serves the purposes of improving the service processes and customer' satisfaction. Furthermore, the study on the perceived quality of online reviews revealed how this area impacts the use of user generated contents into the marketer's interactivity strategies[15]. There has also been an examination made on viral marketing especially targeting the Gen-Y in India and with assessments on the efficiency of viral campaigns heavily relying on eWOM[16].

B. Online Shopping popularity

The rise of online shopping has reshaped the buying process, changing how brands connect with their target audiences. These online reviews possess a unique ability to influence thoughts and behaviors through genuine and relatable content. Research highlights their potential to boost productivity by offering motivation and practical advice to other customers. For example, reviews on websites like Amazon, BigBasket, and Flipkart in niches like health, and wellness, encourage another customer to adopt new healthy habits and develop their skills[5]. By opening up about their experiences and sharing their opinions.

One of the studies examines the persuasive influence of social media personalities through the application of the elaboration likelihood model of persuasion. Specifically, the study finds that when followers perceive influencers as credible, trustworthy, and relatable, they are more likely to purchase products or services promoted by the influencer[13]. We need to focus on the buyer-seller relationship based on the characteristics of influencers like homophily, expertise, trustworthiness, credibility, product relatedness, entertainment value, informative value, and physical attractiveness concerning customer engagement and purchase intention. It identified that such characteristics are directly linked with consumer attention to offers made by influencer and their willingness to purchase[14].

Influencer marketing has emerged as an effective form of online advertising implemented instead of conventional TV commercials, for example, using such influencers as Instagram and YouTube stars. Influencers are usually determined by people's popularity within a certain field of interest based on the graph-based methodology commonly used by researchers. According to the data, mega and mini influences are distinguished by a specific rate of engagement with users' content. These influencers are then used to promote brands, increase reach, and help bolster business growth[15]. To critically assess the role and importance of digital influencers, it's important to look at the background, growth, key factors, and effects of influencer advertising. Additionally, an online questionnaire survey for internet users can be conducted to evaluate how effective using digital influencers can be[16]. Another survey conducted on a sample of Instagram posts revealed that Micro-influencers are more active, creating commercial content more often, and that the use of the product is aligned in association with lifestyle and social events. While commercial ones are less frequent, the presentations are less creative and seem to directly address the audience by influencers with a big number of subscribers. This research offers an empirical contribution to the limited literature on the commercial nature of content posted by social media influencers[17].

III. METHODOLOGY

The following section focuses on the methodology and the current work undertaken, as outlined in this study.

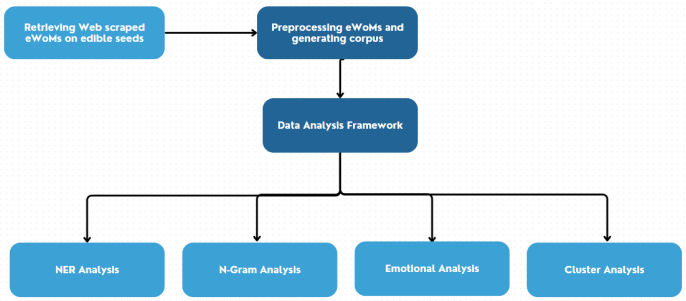

Fig. 1. Analytical flowchart of the present study

*A. **Data Collection***

The first step of the project involved gathering social media data related to seeds from various platforms such as Amazon, BigBasket, and Flipkart. This included collecting user comments, reviews, replies, and other forms of audience interaction. The data captured key elements such as the text of the comments and posts, along with engagement metrics like upvotes and replies. These interactions provided a rich source of information for further analysis, giving us a broad view of how audiences engage with and react to seeds across different E-commerce platforms.

TABLE I
DESCRIPTION OF DATA COLLECTED

Source	Youtube,Amazon, BigBasket & Flipkart
Type	Textual
Instances	750
Missing values?	No
Labelled?	Yes
Duration	4th July 2023 - 18th Sep 2024

B. *Data Preprocessing*

So, after gathering the necessary data, the next step toward proper preparation for analysis is the preprocessing of eWOMs retrieved to get rid of irrelevant content before analysis. In this perspective English stopwords are eliminated, followed by the removal of digits, special symbols, and punctuation symbols.

Cleaning the raw data to remove unwanted elements such as punctuation, special characters, stop words, or URLs since these would bring in no useful information to use for the analysis. After this process, clean data resulted in tokenization: splitting the text into individual words or tokens. This was also a critical step for text preparation for tasks that were to be performed further, such as sentiment analysis and topic modeling, ensuring the text is consistent and in a manageable format for machine learning algorithms.

TABLE II
SAMPLE OF EWOMS DATA COLLECTED

Source	Comment	Label
Bigbasket	good quality	+ve
	become black, bad smell, bad quality	-ve
	superfood chia	Neutral
Amazon	good	+ve
	bad product	-ve
	lesser quantity mentioned sent	Neutral
FlipKart	good product, nicely packed, fresh taste good	+ve
	don't like it's salty	-ve
	tastier	Neutral
YouTube	like it	+ve
	tasteless	-ve
	sunflower seed started hooked eating every day	Neutral

- **Converting to Lowercase:** In methods of preprocessing of the dataset all of the documents to the lowercase. This means that even "Great" and "great" are considered the same word.
- **Tokenization:** This process entails splitting a comment into some smaller units of analysis called tokens which are mere words. In this analysis, these tokens are employed for further analysis or feature extraction of the sentiment analysis model.
- **Stop Word Elimination:** Most common stop words include "is", "the", "a", "an", and "in" that relate to standard meanings that would not possess any or much value information. This process lessens the extensive volume of data that the modeling engine of the learning platform must wade through bringing focus to the more valuable words.

C. *Emotional Analysis*

The study of emotions to classify the sentiments of consumers on edible seeds by employing machine learning algorithms. The emotional analysis categorizes feelings toward edible seeds into three distinct classes: Positive (3) Negative (1) and Neutral (2). Positive Self Reported Emotions mean that people are satisfied when using these seeds to boost their health while Neutral Self Reported Emotions mean that people have nothing in particular against using seeds to boost their health. Self-generated negative affect refers to the parts of the affective experience that are associated with displeasure. It is important to know these emotional responses because customers with positive attitudes toward a product tend to buy the product again. To conduct this analysis, the study employs the 'syuzhet' module in R programming, which identifies and categorizes emotions into eight types: positive and negative emotions entail faith, pleasure, hope, novelty, apprehension, rage, and distaste. This approach permits to analysis of the clients ' behavior in the peculiar context of the edible seeds, which can be found in the online texts. The applications arising from this research can be seen in refining how marketing techniques and products match consumer's emotions. With the help of the results of the emotional analysis, businesses may be more efficient in predicting and satisfying customer demands, customer satisfaction may grow. [17].

D. *Named Entity Recognition*

Further entities in the preprocessed text were found using NER's major natural language processing technique. It was applied to extract and categorize important entities such as goods, products, or individual seeds in an online shopping review. This improves the analysis because it brings structured insights into the seeds discussed and how they are connected to health. By identifying these entities, we were able to gain a clearer understanding of the key components driving user engagement.

E. *N-gram Analysis*

N-gram is a method in text analysis where to analyze sequences of words in the corpus, which can be applied, for instance, to the analysis of consumer's attitudes towards edible seeds. This method analyses texts into bi-grams (occurrences of two words) and tri-grams (occurrences of three words and their associations in consumer reviews. By analyzing n-gram of the feedback collected from the World Wide Web, we can determine useful phrases concerning the virtues or shortcomings of edible seeds. With this knowledge, companies can adjust their advertising techniques while updating their goods with data about peoples' attitudes.

F. *Text Clustering*

To analyze further the online review data, we also made use of text clustering to cluster similar comments. With text clustering, we were able to discover patterns by grouping texts with common themes or sentiments. For this purpose, we made use of Agglomerative Clustering-where every comment is initially treated as its cluster and then these are gradually merged based on their similarities. This method assisted us in making a start small, rather with more focused groupings, and working up to larger, more significant clusters. It provided a much clearer view of how different comments or discussions are related.

We applied Hierarchical Clustering, which resulted in a tree-like structure or dendrogram that shows how comments were grouped at different levels of similarity. This allowed us to intuitively understand the relationships among clusters and see how closely topics and sentiments were related. Taken together, we were able to use these types of clustering approaches to organize data in a much more accessible way to analyze audience reactions to seeds.[18]

IV. Results

We have interesting insights in our study on the Online shopping reviews conversations about seeds and their pros and cons. Using data sources coming from platforms like Amazon, BigBasket, YouTube, and Flipkart, we found the words that appear most often and plotted them using a bar chart and a word cloud. The word cloud gave us a simple yet clear picture of the key topics that were being discussed. We then applied Agglomerative and Hierarchical Clustering to cluster similar posts and hence to be able to see trends and relationships between the discussions. Dendrograms produced from these methods helped us understand how topics relate to each other and how a conversation evolves.

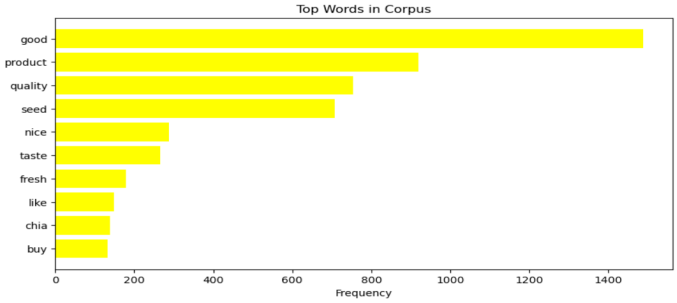

Fig. 2. Top Words in Corpus

Fig. 3. Word Cloud

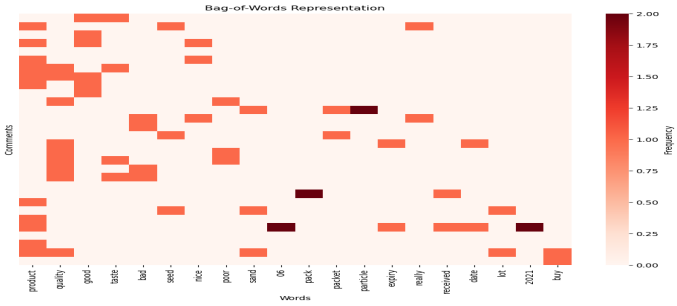

Fig. 4. BoW identified from text corpus

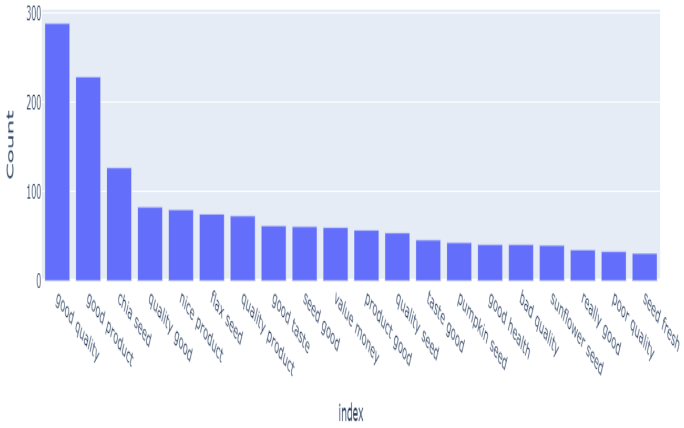

Fig. 5. Outcomes from Bigram analysis

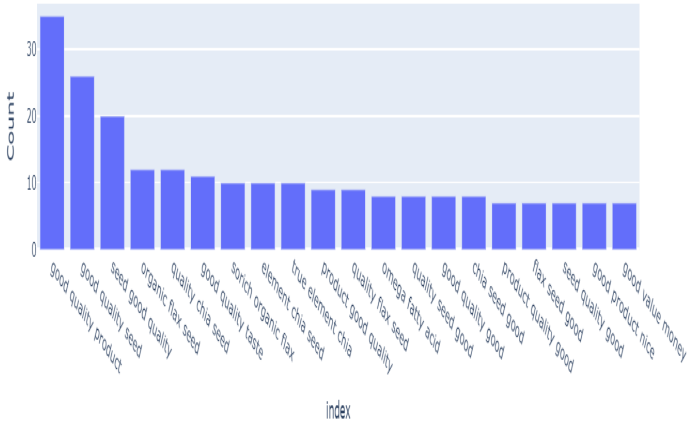

Fig. 6. Outcomes from Trigram analysis

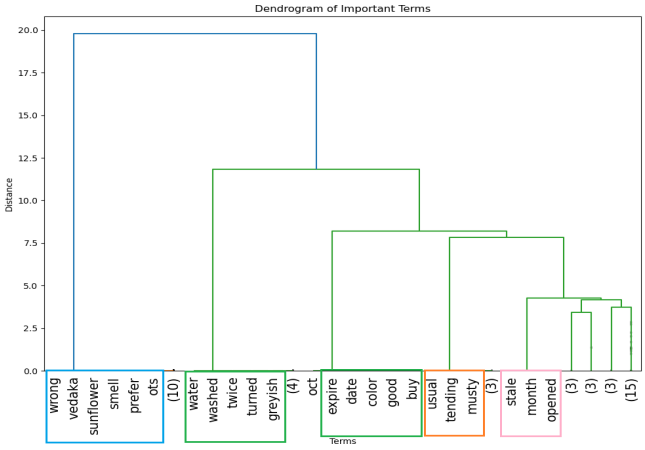

Fig. 7. Dendrogram identified from cluster analysis

V. Conclusion

In this study, we explored how users react on different E-commerce platforms like YouTube, Amazon, BigBasket and Flipkart . By gathering a large dataset and preparing it through preprocessing steps, we were able to organize the raw data into a more structured form. Using Named Entity Recognition (NER), we identified key entities like reviews, products, and the gender of the reviewer , providing a clearer view of the topics while developing newer findings.

We also applied clustering techniques, specifically Agglomerative and Hierarchical Clustering, to group similar posts and comments. This approach allowed us to observe patterns in the discussions, showing how certain topics were related and how conversations shifted over time. Overall, these techniques provided valuable insights into the way people interact with Marques Brownlee's content, helping us better understand audience behavior and engagement across social media platforms.

References

[1] Nandha, R. (2014). Therapeutic potential of sunflower seeds: An overview.

[2] Bashier, I. H., Mosa, M., & Babikir, S. F. (2021). Sesame seed disease detection using image classification. In *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1-5). doi:10.1109/ICCCEEE49695.2021.9429640.

[3] Puraikalan, Y., & Scott, M. (2023). Sunflower seeds (Helianthus annuus) and health benefits: A review. *Recent Progress in Nutrition*, 3(3), 010. doi:10.21926/rpn.2303010.

[4] Savchenko, V., Synyavskiy, O., & Nesvidomin, A. (2022). Pre-sowing treatment of sunflower seeds in a magnetic field. In *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)* (pp. 1-5). doi:10.1109/KhPIWeek57572.2022.9916403.

[5] Savchenko, V., Synyavskiy, O., Zablodskiy, M., Dudnyk, A.,& Nesvidomin, A. (2021). Pre-sowing treatment of flax seeds in a magnetic field. In *2021 IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek)* (pp. 521-526). doi:10.1109/KhPIWeek53812.2021.9570101.

[6] Abbas, S., Sharif, M. K., Sibt-e-Abbas, M., Teferra, T., Sultan, M. T., Anwar, M. J., & Junaid, M. (2022). Nutritional and therapeutic potential of sesame seeds. *Journal of Food Quality*, 6163753.

[7] Bernacchia, R., Preti, R., & Vinci, G. (2014). Chemical composition and health benefits of flaxseed. *Austin Journal of Nutrition & Food Science*, 2(1), 1045.

[8] Mostashari, P., & Mousavi Khaneghah, A. (2024). Sesame seeds: A nutrient-rich superfood. *Foods*, 13(8), 1153. doi:10.3390/foods13081153.

[9] Katare, C., Saxena, S., Agrawal, D., Prasad, G.B.K.S., & Bisen, P.K. (2012). Flax seed: A potential medicinal food. *Journal of Nutrition & Food Science*, 2(1), 10-4172/2155-9600.

[10] Šamec, D., Loizzo, M.R., Gortzi, O., Çankaya, İ.T., Tundis, R., Suntar, I., Shirooie, S., Zengin, G., Devkota, H.P., Reboredo-Rodríguez P., Hassan S.T.S., Manayi A., Kashani H.R.K., & Nabavi S.M. (2022). The potential of pumpkin seed oil as a functional food—A comprehensive review of chemical composition, health benefits and safety. *Comprehensive Reviews in Food Science and Food Safety*, 21(5), 4422-4446.

[11] 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, 2024.

[12] 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, 2022.

[13] S. Gunasekar, P. Das, S. K. Dixit, S. Mandal, and S. R. Mehta, "Wine-experienscape and tourist satisfaction: through the lens of online reviews," Journal of Foodservice Business Research, vol. 25, no. 6, pp. 684–701, 2021. https://doi.org/10.1080/15378020.2021.2006039.

[14] 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), Chennai, India, 2019.

[15] 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), Jaipur, India, 2016.

[16] 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).

[17] Matthew Oct. Jockers, 25, 2024. "Package 'syuzhet'," [Online].2023.Available:project.org/web/packages/syuzhet/syuzhet.pdf

[18] B. Biggio et al., "Poisoning Complete-Linkage Hierarchical Clustering,"2014, pp. 42–52. doi: 10.1007/978-3-662-44415-3 5