A Case Study on Data-Driven Insights into Emotional Responses in Digital Content

Govindu L V N Sridhar, M. Janakiram, M. Vishnu Ram Sai, B. Uma Maheswari, Amulyashree Sridhar
Department of Computer Science and Engineering
Amrita School Of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India,
bl.en.u4cse22117@bl.students.amrita.edu, bl.en.u4cse22129@bl.students.amrita.edu, bl.en.u4cse22128@bl.students.amrita.edu
b_uma@blr.amrita.edu, s_amulyashree@blr.amrita.edu

***Abstract*—Popularity of influencers is largely dependent on consumer opinions. The current study identifies the popularity of an influencer in the educational sector, i.e. 3Blue1Brown, by retrieving learner's opinions from popular social media websites. A data analysis framework is proposed to derive key insights after data collection. In this perspective, emotional analysis, NER analysis, *N*-gram analysis, and cluster analysis are performed. Outcomes from these approaches revealed positive emotions from learners with remarkable prominence given to the linear algebra course offered by the influencer. These outcomes can be utilised by influencers to devise relevant learning content which can further enhance the overall learner's experience.**

Keywords- Influencer, 3Blue1Brown, Emotional analysis, Cluster analysis

I. Introduction

Over the last couple of years, influencers have been the most dominating social media functionalities, especially on Instagram, YouTube, and TikTok. Influencer marketing means an influencer is considered to be an authority or enthusiast in one area, such as beauty, fitness, travel, or technology, and has successfully attracted huge followers for his content. They could directly connect with their audience, which created immense marketing avenues for brands, and hence, influencer marketing. Specifically, as the belief in personal recommendations overshadowed traditional advertisement, these influencers dictate trends; they influence purchasing decisions and receive cultural icon status in the digital world. This is because their popularity is founded on perceived authenticity along with the massive communities built around their brands.

The author talks about the impact of eWOMs on consumer's purchasing decision [1].Identified impact of eWOMs in spreading information on youtube [2].Another study discusses information dissemination through digital platforms using eWOMs and compares it with traditional communication.Outcomes revealed that eWOMs offered a better medium for communication due to accessibility and ease of information exchange without the influence of time [3].Another study discusses consumer's preferences towards academic insights on relevant subjects like mathematics, science, or technology, which will help enthusiastic learners [4]. According to the study, the author discusses the impact of influencers on social media platforms to perceive opinions on diverse domains including learning tools/materials, and other academic products[5].

Influencer credibility is analysed in another study to identify the quality of content, by ensuring that the influencer can increase the visibility and popularity of educational channels [6]. Another study explores the impact of educational influencers towards increasing the marketing strategies of online courses and resources [7].With the emergence of the COVID-19 pandemic, a study explored the reliability of digital educational influencers in guiding learners towards better decision making [8].Another similar study interpreted the impact of educational influencers in post-pandemic era.Outcomes suggested that the influence of digital creators is yet to be examined on a large scale to gain better insights [9].

Based on inferences from previous studies the current study identified the impact of educational influencer, 3Blue1Brown, towards determining learners digital preferences. Additionally, a data analysis framework was employed to analyse knowledge from eWOMs extracted on the influencer. Our contributions in this research work are listed here:

- Collected comments data about the influencer from social media platforms(YouTube, X(Twitter) & Instagram) and claiming the data using tokenization, stop word elimination and converting to lowercase.
- Visualized the collected data using Word Cloud to identify frequently occuring words in comments.
- Performed emotional analysis using
- Categorized the entities collected from the text corpus using NER approach.
- Interpreted the sequence of words using *N*-gram approach.
- Identified the interconnection between the words using hierarchical clustering approach.

II. Related Works

This section highlights prominent studies on influencer popularity analysis.

Initial study explored association between video creators, who run channels regarding coding and analysis. Outcomes revealed Support Vector Machine(SVM) algorithm had an average accuracy of 77% in identifying useful comments. In addition, SumBasic summarization method captured the primary concerns of viewers effectively [10]. Another study identified primary perceptions of Head Mounted Displays(HMDs) of 379 YouTube comments.Outcomes identified that most users had shown positive attitude towards HMDs.Here, negative attitudes were less frequent which often were related to assessment and performance comparision with other products [11].

Another study identified learner's sentiments towards YouTube comments. Here, NLP techniques including VADER lexicon and YouTube API were used to analyze emotions. Results indicated that positive comments were predominant in the data [12]. Another study identified emotions from Arabian Twitter data. Numerous algorithms including k-NN(k-Nearest Neighbors), Naïve Bayes, and SVM were employed to classify the emotions [13]. Another similar study surveys on sentiment analysis approaches towards video recommendation systems on YouTube by comparing previous traditional approaches based on metadata that considers user sentiments [14].

Another study explored the analysis of code-mixed languages, one such flavor being Tanglish a mixture of Tamil and English.The current study used lexicon based and machine learning models to classify sentiments [15]. Another similar study performs sentiment analysis of the YouTube comments by employing machine learning techniques, such as SVM, Naïve Bayes(NB), and k-NN [16].

Another indistinguishable study investigates extent of emotions using machine learning techniques like NB and SVM. This research proved that combined classifiers perform better in sentiment categorization for large data [17]. Another similar study performed literature review of popular text mining approaches towards extraction of knowledge frrom text data. Here, tokenization, stemming, and n-gram analysis was also performed [18].Another study identified survival factors for family-run home stays using text mining approaches. They found some salient features including customer satisfaction and adaptability, which keep such businesses peaceless [19].

Another study identified evolution of consumer behavior,based on demographic changes in modern technologies [20]. Another similar study explored the tourist contentment in wine tourism based on online reviews. Results indicated that the observations are critical to understand customer intentions in E-travel sector [21]. Another similar study identified importance of online feedback towards innovation and satisfaction [22]. On the same lines, another study explored user-generated content has an impact on consumer decisions [23].

Another study identified the impact of viral marketing and brand promotions on purchase intentions of Gen-Y population [24]. Another study examined the subject of sentiment analysis on Facebook by combining lexicon-based and machine-learning approaches to include user emotions. This analysis supports the adaptive e-learning with personalization toward emotional states [25]. Another study explored the extent of rational satisfaction towards services and quality of products [26].

III. Methodology

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

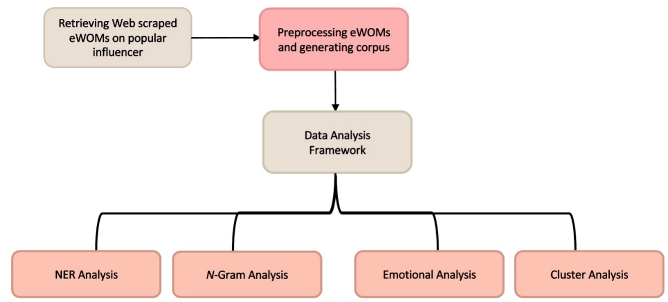

Fig. 1. Analytical flowchart of the present study

A. *Data Collection*

1) Web Scrapping: Here, eWOMs are retrieved from popular social media platforms like YouTube, X(Twitter) and Instagram on learner's opinion towards 3Blue1Brown.These opinions are web scrapped using Python 3.9 as a programming language. Here, modules Selenium and Beutifulsoup were used to retrieve the opinions from above-mentioned resources. The eWOMs are retrieved for a predefined time slot of 15 days to avoid complexities in data. The detailed description of the data retrieved 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	X(Twitter),Instagram, YouTube	No	Yes	1st Aug 2024 - 15th Aug 2024

B. *Data Preprocessing*

The next step involves preprocessing of eWOMs retreived to get rid of irrelevant content prior to analysis. In this perspective English stopwords are eliminated, followed by the removal of digits, special symbols, and punctuation symbols.

TABLE II
SAMPLE OF EWOMS DATA COLLECTED

Sl.No	Comment	Label	Data Source
1	Amazing video as usual	Positive	Youtube
2	Hi, your videos are excellent, Thanks a lot	Positive	Youtube
3	Great content, very elaborative	Positive	X(Twitter)
4	I don't understand, but kudos for the editing	Neutral	Instagram
5	I don't study calculus, but still love it	Neutral	X(Twitter)
6	Animations went too hard for this video, which creates trouble in understanding	Negative	Instagram

Furthermore, the data is converted into lowercase, followed by tokenization and stemming. The preprocessed data is termed as 'corpus' for further analysis.

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 data is further subjected to emotional analysis to identify the extent of contentedness or discontentedness towards the influencer. In this perspective, positive emotions reflect satisfaction towards the content generated by 3Blue1Brown, while neutral emotions reflect no specific opinions towards the influencer. On the contrary negative emotions indicate dissatisfaction with the content generated by the influencer. Presently this study implements supervised classification of emotions using machine learning algorithms.Using this technique, 3 categories of classifications were determined: 3 = Positive; 2 = Neutral; and 1 = Negative. As emotionally satisfied learners are likely to prefer content from the same influencer, it is thereby significant to understand emotions in online reviews. In this perspective, the present study employs the 'syuzhet' module in R programming language to perform emotional analysis. This approach identifies emotions and groups them into 8 categories like trust, joy, anticipation, surprise, fear, anger, and disgust [27].

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

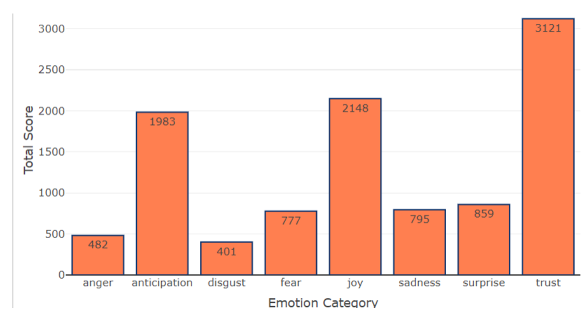

Fig. 2. Emotion Analysis

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.

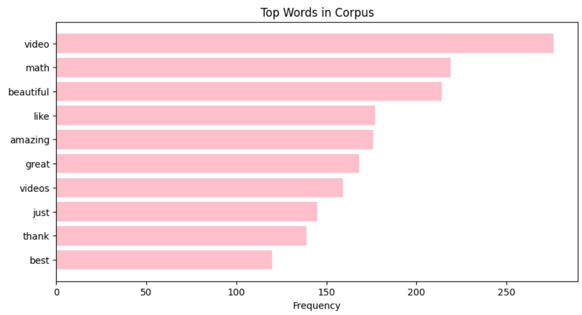

Fig. 3. Top 10 words identified from the corpus

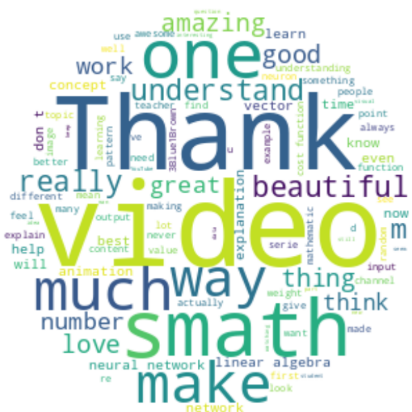

Fig. 4. Word cloud identified from data

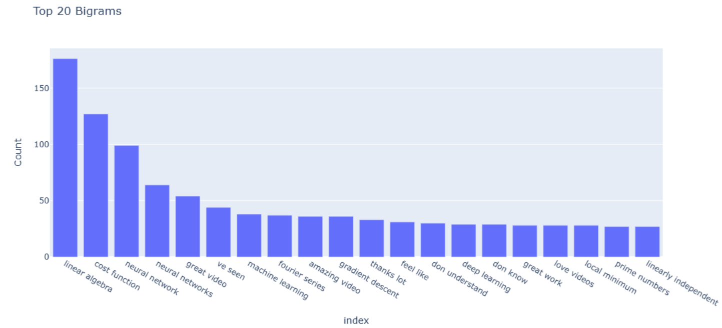

Fig. 5. Outcomes from Bigram analysis

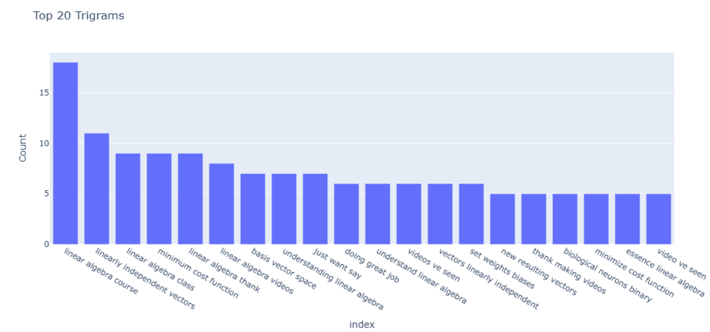

Fig. 6. Outcomes from Trigram analysis

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.

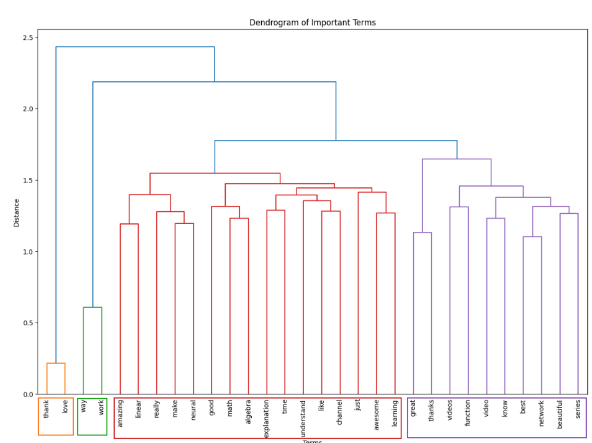

Fig. 7. Dendrogram identified from cluster analysis

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	Frequency	Category of Entity
Like	434	Verb
Thank	501	Verb
Amazing	315	Adjective
Math	355	Noun
Video	512	Noun
Understand	328	Verb
Great	338	Adjective
Much	308	Adjective/Adverb

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

V. Conclusion

The current study retrieved learner's opinions towards a popular influencer i.e. 3Blue1Brown and conducted a structured analysis based on emotions, NER, *N*-gram, and clustering. Outcomes revealed that learners had optimistic impressions towards the influencer as terms like "Thank", "beautiful", "amazing", and "understand" were observed in the text corpus. These worthwhile sentiments can be attributed due to the liking of mathematical course i.e. linear algebra. Additionally, negative sentiments were negligible in the corpus. Also, cluster analysis confirmed the outcomes by identifying productive opinions towards the influencer.

However, there were few limitations in the current study. Primarily, the data collected was restricted to a particular influencer, hence the sample was homogeneous. Thereby, generalization of outcomes is not feasible to other influencers. Also, the data was collected for a pre-defined time, which restricted the scope for advanced analysis. In this context, implementing a comprehensive analysis on the influencer could reveal finer insights on the emotions of learners.

References

[1] Dellarocas, Chrysanthos. (2011). The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. Management Science. 49. 10.1287/mnsc.49.10.1407.17308.

[2] Cheung, Christy & Thadani, Dimple. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. Decision Support Systems. 54. 461–470. 10.1016/j.dss.2012.06.008.

[3] Huete-Alcocer, Nuria. (2017). A Literature Review of Word of Mouth and Electronic Word of Mouth: Implications for Consumer Behavior. Frontiers in Psychology. 8. 10.3389/fpsyg.2017.01256.

[4] Chivandi, Abigail & Samuel, Olorunjuwon & Muchie, Mammo. (2018). Social Media, Consumer Behavior, and Service Marketing. 10.5772/intechopen.85406.

[5] Erkan, Ismail & Evans, Chris. (2016). Social media or shopping websites? The influence of eWOM on consumers' online purchase intentions. Journal of Marketing Communications. 1-17. 10.1080/13527266.2016.1184706.

[6] Khwaja, Muddasar & Zaman, Umer. (2020). Configuring the Evolving Role of eWOM on the Consumers Information Adoption. Journal of Open Innovation Technology Market and Complexity. 6. 1-13. 10.3390/joitmc6040125.

[7] Bashir, Shahid & Khwaja, Muddasar & Mahmood, Asif & Ali, Jamshid & Latif, Fawad. (2021). Refining e-shoppers' perceived risks: Development and validation of new measurement scale. Journal of Retailing and Consumer Services. 58. 102285. 10.1016/j.jretconser.2020.102285.

[8] Hall, Colin & Prayag, Girish & Fieger, Peter & Dyason, David. (2020). Beyond panic buying: consumption displacement and COVID-19. Journal of Service Management. 10.1108/JOSM-05-2020-0151.

[9] Le-Hoang, Phuong Viet. (2020). The effects of Electronic Word of Mouth (eWOM) on the adoption of consumer eWOM information. Independent Journal of Management & Production. 11. 1749-1766. 10.14807/ijmp.v11i6.1152.

[10] E. Poché, N. Jha, G. Williams, J. Staten, M. Vesper and A. Mahmoud, "Analyzing User Comments on YouTube Coding Tutorial Videos," 2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC), Buenos Aires, Argentina, 2017, pp. 196-206, doi: 10.1109/ICPC.2017.26.

[11] N. Zuidhof, S. B. Allouch, O. Peters and P. -P. Verbeek, "Anticipated Acceptance of Head Mounted Displays: a content analysis of YouTube comments," 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan, 2019, pp. 399-402, doi: 10.1109/PERCOMW.2019.8730658.

[12] R. Pradhan, "Extracting Sentiments from YouTube Comments," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 1-4, doi: 10.1109/ICIIP53038.2021.9702561.

[13] W. W. Thant, S. Khaing and E. E. Mon, "Preprocessing of YouTube Myanmar Music Comments for Sentiment Analysis," 2020 IEEE Conference on Computer Applications(ICCA), Yangon, Myanmar, 2020, pp. 1-6, doi: 10.1109/ICCA49400.2020.9022834.

[14] H. Bhuiyan, J. Ara, R. Bardhan and M. R. Islam, "Retrieving YouTube video by sentiment analysis on user comment," 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, 2017, pp. 474-478, doi: 10.1109/ICSIPA.2017.8120658.

[15] M. Janotheepan, S. D. H. S. Wickramarathna, M. J. A. Amas, T. Rajeetha, A. K. L. M. Farhath and H. A. F. Sanas, "Sentiment Analysis for YouTube Cooking Recipes Videos Using User Comments," 2024 4th International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2024, pp. 235-240, doi: 10.1109/ICARC61713.2024.10499736.

[16] R. F. Alhujaili and W. M. S. Yafooz, "Sentiment Analysis for Youtube Videos with User Comments: Review," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 814-820, doi: 10.1109/ICAIS50930.2021.9396049.

[17] Muhammad, Abbi & Bukhori, Saiful & Pandunata, Priza. (2019). Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier. 199-205. 10.1109/ICOMITEE.2019.8920923.

[18] S. Naik and A. Katre, "YouTube Universe of Comments: A Machine Learning approach for systematic classification of YouTube Comments on custom prepared dataset," 2023 World Conference on Communication & Computing (WCONF), RAIPUR, India, 2023, pp. 1-5, doi: 10.1109/WCONF58270.2023.10235049.

[19] 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.

[20] 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.

[21] 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, Nov. 2022, doi: 10.1080/15378020.2021.2006039.

[22] 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.

[23] 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.

[24] 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.

[25] Alvaro Ortigosa, José M. Martín, and Rosa M. Carro. 2014. Sentiment analysis in Facebook and its application to e-learning. Comput. Hum. Behav. 31, C (Feb 2014), 527–541. https://doi.org/10.1016/j.chb.2013.05.024

[26] Yu, Ting & Dean, Alison. (2001). The contribution of emotional satisfaction to consumer loyalty. International Journal of Service Industry Management. 12. 234-250. 10.1108/09564230110393239.

[27] Matthew Jockers, "Package 'syuzhet'," 2023. Accessed: Oct. 25, 2024. [Online]. Available: https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf

[28] B. Biggio et al., "Poisoning Complete-Linkage Hierarchical Clustering," 2014, pp. 42–52. doi: 10.1007/978-3-662-44415-3_5.