Popularity Analysis of Social Media Influencer: A Data Centric Approach

Amulyashree S.
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
s_amulyashree@blr.amrita.edu

Ruchi Chaurasiya
Department of computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22278@bl.students.amrita.edu

Shakshi Yadav
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22286@bl.students.amrita.edu

Swagata Malik
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22252@bl.students.amrita.edu

***Abstract*—The success of content-driven platforms like The Ranveer Podcast Show lies in audience engagement fueled by electronic Word of Mouth (eWOM). Influencers like Ranveer Allahbadia shape public opinion across personal growth, fitness, and entrepreneurship by creating authentic, relatable content that builds trust and surpasses traditional media in influence.**

This study explores audience perceptions of The Ranveer Show through sentiment analysis, N-gram analysis, and cluster analysis, revealing positive viewer sentiments and appreciation for its value-driven discussions. As India's Influencer economy grows with platforms like YouTube, the podcast exemplifies the transformative power of eWOM in shaping public discourse and consumer behavior.

***Index Terms*—Social Media Influencer, eWOM,Influencer , sentiment analysis, N-gram analysis, cluster analysis**

I. Introduction

An influencer is someone who uses social media to shape opinions, behaviors, and purchasing decisions by sharing relatable and engaging cont. Their popularity stems from building trust through authentic connections with their audience, often surpassing the influence of traditional media. Globally, countries like the United States lead in the number and impact of influencers, while India is quickly catching up. The rise of platforms like YouTube and Instagram, combined with affordable internet access and high data consumption, has created fertile ground for influencers to thrive in diverse industries such as fashion, fitness, and education.Below are some notable research contributions in this field. [1]The study explained the theory of planned behavior that deals with the various aspects of marketing tactics performed by different social media influencers. [2]The research addressed the influence of digital video promotions and social media influencers on purchasing of fitness and health related products.

One driving factor behind this growth is Electronic Word of Mouth (eWOM), where users share recommendations online, amplifying influencers' reach and credibility. Social media users increasingly rely on peer opinions and reviews, creating a feedback loop that further enhances influencers' popularity.

A prominent example in India is Ranveer Allahbadia, host of The Ranveer Show, a podcast addressing topics like entrepreneurship, personal growth, fitness, and mental health. Ranveer's authentic style and ability to connect with his audience through inspiring conversations have positioned him as a leading influencer. His success is driven by consistent engagement, meaningful discussions with experts, and leveraging eWOM to expand his reach.

This report explores the growth of The Ranveer Show, analyzing audience metrics, engagement rates, and the impact of eWOM. By understanding the factors contributing to Ranveer's success, the report sheds light on the broader trends in influencer marketing and content creation in India.

II. Literature Survey

In this digital era, most of the people started depending on social media influencers for the reviews of products or brands. Moreover, Some social media influencers have caused a lot of positive as well as negative impacts on today's generation people. Some of the related works done in this field are explained below.

[3] The research introduced the literature reviews of many social media influencers that shed lights on different topics collaborating with audiences where most of the topics are still unexplored. It also explained how influencers has less impact on Indain consumers compared to other continents and country.

[4] The research discussed how social media influencers build professionals relations with brand sponsors and advertise products for promotions. It also signifies the astonishing uptick of influencers marketing compared to traditional paid promotions.

[5]The reseach investigated the dimensions of wine-experienscape that influence customer satisfaction in the con-

text of wine tourism. It utilized the framework of experienscape to explore various dimensions relevant to wine tourism experiences.The study concluded with implications for both managerial practices and theoretical frameworks in the field of wine tourism.

[6]The study looked at the impacts of two types of celebrities, Instagram celebrities and conventional celebrities, on source dependability, brand demeanor, envy, and social proximity.

[7]The study proposed research work on how social media influencers make a great impact on customers buying behaviour and also how the gender plays a vital role in shopping. Most, Zen G customers are influenced by the product/brand promotions done by many social media influencers.

[8]The study explored the social responsibilities, loyalty and psychological well-being of the viewers/customers towards their influencers.

[9]The research outlined a comprehensive study that leveraged text mining techniques to analyze user reviews of homestays, aiming to provide valuable insights for improving service quality and marketing strategies in the hospitality sector.

[10]The outline employed bibliometric analysis, which involved analyzing bibliographic records from Scopus. This method allowed researchers to quantitatively assess growth and trends in consumer behavior research published in the Journal of Consumer Behavior.

[11] The study introduced mixed method analysis of the content on data collected from 151 political leader interviews taken on youtube channels. It was found that the content delivered through youtube channels had slowly taken over old set-up-box television news

[12]The study looked at how customer reviews affect innovations in e-travel providers' service processes. The study focused on online reviews that are posted on blogs and social media sites, exposing unfavorable opinions and notable areas of discontent.

III. Implementation

This analysis explores sentiment analysis and topic modeling on comments from The Ranveer Show, sourced from YouTube and Reddit.

A. Sentiment Analysis

1. Setup: We use 'pandas' for data handling and 'nltk' for text processing, including tokenization, stopword removal, and lemmatization. 2. Preprocessing: Comments are cleaned to remove noise and prepare for analysis. 3. Scoring: Using 'SentimentIntensityAnalyzer', each comment receives a compound score:

- Positive (0.05): Label 1
- Negative (-0.05): Label -1
- Neutral (-0.05 to 0.05): Label 0

Neutral comments are excluded for binary sentiment analysis.

B. Topic Modeling

1. Preparation: Text is processed by removing stopwords, numbers, and punctuation, and terms with low document frequency are filtered out.

2. Insights: - Top Words Themes: Highlighted common themes such as cultural heritage and personal development. - Topic Relationships: Visualized correlations between topics and key themes. - Word Clouds: Visualized significant terms for specific topics. - Proportions: Identified the most discussed topics within the dataset.

Combining sentiment analysis and topic modeling uncovers the general tone of audience feedback and specific themes driving discussions, providing actionable insights to shape future content strategies.

IV. Methodology

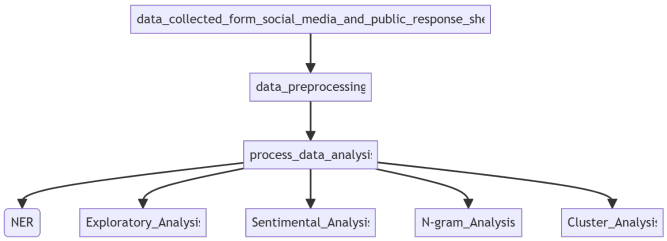

Fig. 1. Framework of Proposed Work

The following section describes the methodology of the proposed work as visualized

A. Data Collection

In this project, electronic word-of-mouth (eWOM) data is collected to analyze public opinions on The Ranveer Show podcast, hosted by Ranveer Allahbadia on the BeerBiceps YouTube channel. The eWOM data is sourced from popular platforms such as YouTube, X (formerly Twitter), and Instagram, with a fixed 15-day time frame to maintain consistency and reduce complexity.

Types of Data Collected:

Social Media : Direct comments from platforms like YouTube, X, and Instagram, specifically on the BeerBiceps channel.

Public Response Data: Feedback collected via a custom Google Form, allowing participants to share their opinions on the podcast.

TABLE I
DESCRIPTION O DATA COLLECTED

Type	Textual
Data Instances	2150
Data source	X(Twitter),Instagram,Facebook,You Tube
Missing values?	No
Label	Yes
Duration of Data Collection	1st Aug 2024- 15th Aug 2024

The collected data is compiled into an Excel sheet for easy organization and analysis. This method avoids the need for

Python-based web scraping, ensuring a clear and straightforward representation of data from all sources.

B. Data Preprocessing

Irrelevant elements, such as stopwords, digits, special characters, and punctuation, are removed to streamline the text data for analysis.

TABLE II
SAMPLE OF E WOMS DATA COLLECTED

SL.NO	Comment	Label	Data Source
1	Very informative podcast	Positive	You Tube
2	Amazing podcast	Positive	You Tube
3	so inspiring	Positive	Instagram
4	wish to go	Neutral	Facebook
5	find the killer	Neutral	Facebook
6	too egoistic	Negative	Instagram

C. Exploratory Analysis

To extract significant terms from the text corpus, the study employs Term Frequency (TF) and the Bag-of-Words (BoW) model. This process identifies frequently occurring words within the corpus. Term Frequency (TF) is defined mathematically as:

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

Equation 1: Term Frequency Formula

Here, x_a represents a specific term, and y_b refers to a document within the corpus.

Additionally, the Bag-of-Words (BoW) method is applied, treating terms as elements of a "bag." This approach transforms the text data into fixed-length numerical vectors, enabling deeper analysis of relevant terms in the eWOM data.

D. Named Entity Recognition (NER)

Named Entity Recognition (NER) is a common text processing technique that involves identifying and classifying entities into predefined categories. In this study, NER is applied to the text corpus to detect and categorize various entities.

E. Emotional Analysis

Emotions in the data are classified into positive (1), neutral (0), and negative (-1) categories using the syuzhet package in R. Emotions are further categorized into trust, joy, anticipation, surprise, fear, anger, and disgust to gauge public sentiment towards the content.

F. N-gram Analysis

Bi-grams and tri-grams reveal word associations and patterns, providing insight into common co-occurrences in the text.

G. Text Clustering

Hierarchical clustering with a Euclidean distance matrix and complete linkage method groups terms to explore relationships and connections within the corpus.

V. Result and Conclusion

A. Name Entity Rcognition(NER)

From social media dataset , A Viewer comments on the Ranveer Show reveal strong positive feedback, appreciating guest diversity, emotional resonance, and personal growth themes. YouTube is a key engagement platform, with a loyal audience valuing authentic discussions. These insights suggest focusing on diverse, impactful content to further strengthen audience connection.

Fig. 2. Entity Visualization from social media(X,you tube,instagram) dataset

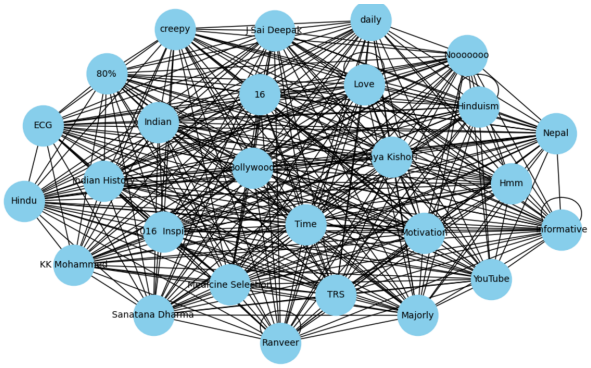

Fig. 3. Entity Visualization for public response dataset

From public response dataet The network graph shows strong connections between themes like "Indian History," "Bollywood," "Hinduism,""Motivation" in viewer comments. The dense links indicate high engagement with topics blending cultural heritage and personal growth.

Fig: 4 word cloud shows that viewers of The Ranveer Show podcast frequently use positive words like "love," "great," "amazing," and "informative" while words like "life," "knowledge," and "inspiration" highlight the value people find in the content. Overall, it reflects a strong appreciation for the show's quality, relatability, and informative nature.

This word cloud shows that viewers of The Ranveer Show podcast frequently describe it as "interesting," "informative," and "amazing." Words like "knowledge," "motivation," and

Fig. 4. Word cloud identified from social media(X,you tube,instagram) dataset

Fig. 5. Word cloud identified from public response dataset

”life” suggest that people value the insightful and varied topics covered in the show.

B. Explatory Analysis

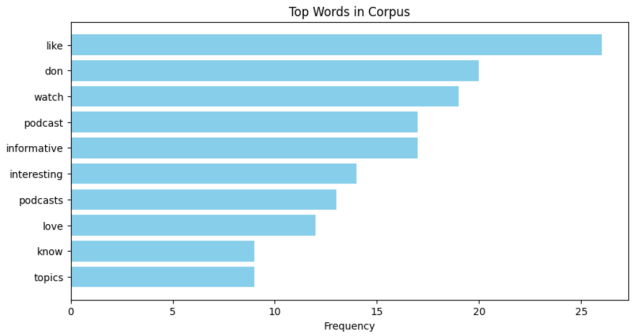

Fig. 6. Top 10 words identified from corpus of public response dataset

Fig : 6 bar chart shows the most frequently used words in comments about ”The Ranveer Show.” Words like ”like,” ”watch,” ”podcast,” and ”informative” are among the most common, suggesting that viewers often discuss their enjoyment of the show, its format, and its educational value.

C. Sentimental Analysis

This sentiment distribution chart shows that most viewers have a positive sentiment towards The Ranveer Show podcast, with a large peak around 1.0, indicating high positive sentiment. There is also a smaller peak near 0, representing

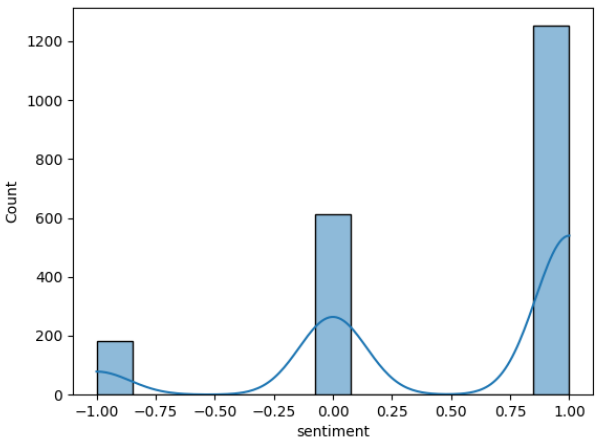

Fig. 7. Sentiment analysis for social media(X,You Tube,Instagram dataset

neutral feedback, and a smaller count around -1.0, showing a minority of negative sentiment. Overall, the sentiment is mostly positive.

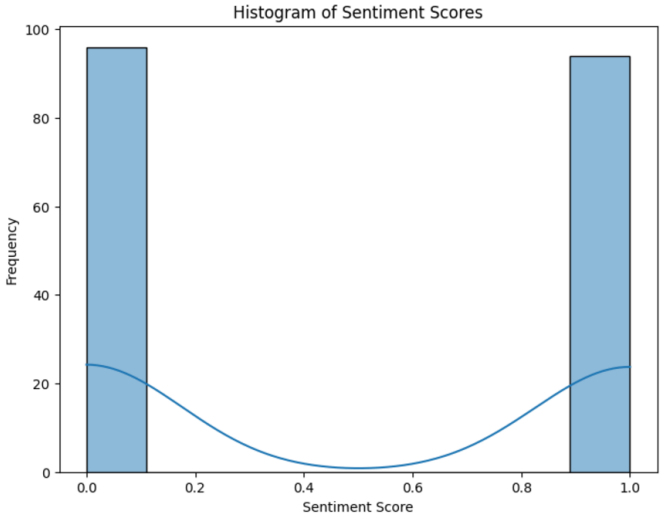

Fig. 8. Sentimental Analysis from public response dataset

For public response dataset histogram shows that comments on the Ranveer Show podcast are highly polarized, with most sentiment scores at 0 (very negative) or 1 (very positive). Few comments fall in the middle, indicating that viewers tend to feel strongly—either very favorable or critical.

D. N-gram Analysis

To uncover associations between terms, experiments were conducted using N-gram analysis. For n=1, no notable associations were found. However, with n=2 (Bigram Analysis), significant patterns like ”informative interesting” and ”guest amazing” emerged, as illustrated in Fig 13.

When analyzing n=3 (Trigram Analysis), relationships such as ”valo lage na” and ”love listening guest” were identified,

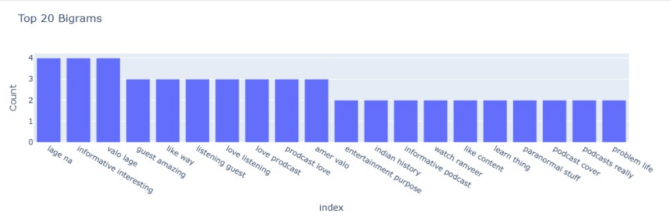

Fig. 9. Outcomes from Bigram analysis from public response dataset

Fig. 10. Outcomes from Trigram Analysis from public response dataset

as depicted in Fig 14. Both figures highlight the top 20 associations, showcasing the key themes present in the comments

E. Cluster Analysis

Cluster Analysis identified key themes in the eWOM data, grouping comments based on similarities in sentiment and content. Positive feedback centered on motivation and personal growth, while negative comments focused on perceived egoism. This clustering provides valuable insights into audience preferences and emotional responses, guiding future content strategies to better engage viewers and address concerns.

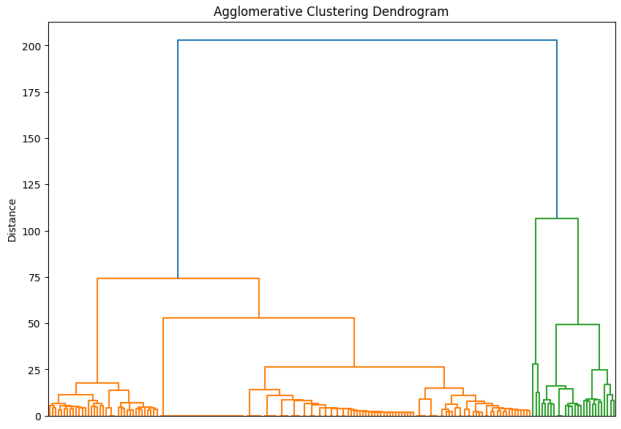

Fig. 11. Dendrogram of Agglomerative Clustering from public response dataset

fig: 11 dendrogram represents agglomerative hierarchical clustering of comments from "The Ranveer Show Podcast." Each branch shows how comments are grouped based on similarity, with the height indicating the distance (or dissimilarity) between clusters.

Fig:12 dendrogram for The Ranveer Show Podcast analysis shows distinct clusters of viewer sentiments, indicating varied

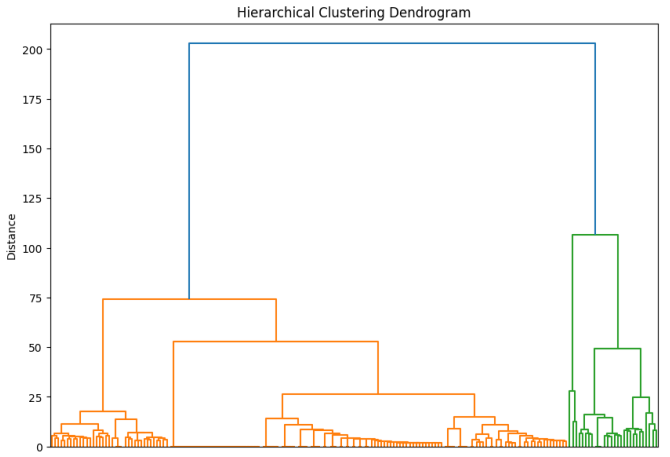

Fig. 12. Dendrogram of Hierarchial Clustering from public response dataset

patterns of positive, neutral, and negative feedback. This clustering highlights key themes and sentiment trends in audience reactions.

VI. Conclusion

This study on The Ranveer Show demonstrates how electronic Word of Mouth (eWOM) drives audience engagement and sentiment analysis. By applying techniques such as sentiment classification, emotional analysis, and topic modeling, we identified positive viewer sentiments and key themes that resonate with the audience. The insights gained can help influencers optimize content strategies and improve audience connection. Overall, the research highlights the growing influence of digital platforms and eWOM in shaping consumer behavior and public opinion in India's influencer-driven economy.

VII. Future Scope

Future research could focus on refining sentiment and emotion detection models to improve accuracy, especially with regard to ambiguous or nuanced comments. Expanding the study to include more influencers or podcasts could provide a broader perspective on eWOM's impact. Additionally, incorporating real-time data analysis and exploring other platforms, such as TikTok, could enhance the understanding of audience dynamics across different social media environments. Improving the scalability of the analysis framework would allow for the study of larger datasets in more diverse contexts.

References

[1] Anjali Chopra, Vrushali Avhad, Jaju, and Sonali. Influencer marketing: An exploratory study to identify antecedents of consumer behavior of millennial. *Business Perspectives and Research*, 9:77–91, 2021.

[2] Suman Devi and SN Mahapatra. Social media influencers: Key to influence consumer's intention to purchase fitness products. *International Journal of Health Sciences*, pages 3181–3188, 2022.

[3] J Dewangan, I Siddiqui, and U Trehan. Social media influencer marketing: A systematic literature review. *International Journal of Business Excellence*, 1:1, 2022.

[4] RM Duffy and BD Kelly. Can the world health organisation's 'qualityrights' initiative help reduce coercive practices in psychiatry in ireland? *Irish journal of psychological medicine*, 40:114–117, 2023.

[5] Sangeetha Gunasekar, Payel Das, Saurabh Kumar Dixit, Santanu Mandal, and Simrran R Mehta. Wine-experienscape and tourist satisfaction: through the lens of online reviews. *Journal of Foodservice Business Research*, 25:684–701, 2022.

[6] S Venus Jin, Aziz Muqaddam, and Ehri Ryu. Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, 37:567–579, 2019.

[7] Nimish Kadam, Anindita Deshmukh, and Rajashri Kadam. A study on impact of social media influencers' endorsements on the buying behaviour of gen z, for lifestyle and electronics product category with special reference to pune city. *Parikalpana: KIIT Journal of Management*, 17:218–227, 2021.

[8] Jihye Kim and Minseong Kim. Rise of social media influencers as a new marketing channel: Focusing on the roles of psychological well-being and perceived social responsibility among consumers. *International journal of environmental research and public health*, 19:2362, 2022.

[9] Jay Krishnan, Biplab Bhattacharjee, Maheshwar Pratap, Janardan Krishna Yadav, and Moinak Maiti. Survival strategies for family-run homestays: analyzing user reviews through text mining. *Data Science and Management*, 7:228–237, 2024.

[10] Weng Marc Lim, Satish Kumar, Nitesh Pandey, Deepak Verma, and Divesh Kumar. Evolution and trends in consumer behaviour: Insights from journal of consumer behaviour. *Journal of Consumer Behaviour*, 22:217–232, 2023.

[11] Rudransh Mukherjee, Shaily Desai, Dhruv Raghavan, and Joyojeet Pal. Political interviewing on youtube: An analysis of indian politicians' engagement with digital influencers and media houses online.

[12] G Rejikumar, V Nimisha, and Shreya 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)*, pages 152–157. IEEE, 2019.