Big Data Analytics Mini Project
on
Effects of Social Media on Indian Culture Using Data Analytics

A Mini-Project Report Submitted
For
Partial Fulfillment of the Requirements
of the Degree of
Bachelor of Engineering

In

COMPUTER ENGINEERING
(Semester VII)

By

Basilica Anthony 9524
Andre John 9544
Sharli Khot 9550

Under the guidance of
Prof. Prachi Patil

DEPARTMENT OF COMPUTER ENGINEERING
Fr. Conceicao Rodrigues College of Engineering
Bandra (W), Mumbai - 400050
University of Mumbai

2024-25

Index

Abstract

This study investigates the effects of social media on Indian culture by analyzing patterns of engagement on popular platforms such as Instagram, Twitter and Facebook. Using a dataset detailing social media usage among diverse demographics in India, we examine key metrics including follower counts, engagement levels, platform usage frequency, and demographic factors like age, education level, and professional status.

Through various data visualization techniques, such as histograms, violin plots, and scatter plots, we identify distinct patterns of social media use across age groups, education levels, and employment statuses. Additionally, by comparing weekday and weekend usage, this study explores differences in leisure behavior and its cultural implications in the Indian context.

Findings highlight generational shifts in communication, differing usage levels based on educational attainment, and varied engagement between working professionals and students, suggesting that social media contributes to evolving social norms and cultural dynamics.

This analysis of digital behavior sheds light on how social media influences identity, information consumption, and social interactions, offering a data-driven perspective on the cultural transformation within India. Through this study, we aim to provide insights into the digital footprints of Indian society and the broader cultural shifts driven by social media engagement.

INTRODUCTION

Social media has become an influential force in shaping cultural practices, communication styles, and social dynamics in India. As platforms like Instagram and Facebook grow in popularity, they not only reflect individual behavior but also collectively impact societal norms and cultural values.

This study aims to explore these impacts by analyzing social media usage patterns among Indian users. Using R for data analysis and visualization, we examine a dataset that includes diverse demographic and social media usage details, such as follower counts, posting frequency, and total time spent on social media.

The R code provided here enables a comprehensive analysis through the lens of various visual and statistical tools. First, we preprocess and filter the data to focus on specific groups, such as working professionals and students. Then, we perform targeted analyses using histograms, probability mass functions, scatter plots, and bar charts to reveal trends in social media engagement across demographic categories, including age, education level, and region.

The goal is to highlight patterns of digital engagement and draw insights into how social media influences cultural aspects like identity, connectivity, and information sharing within India.

By quantifying social media behaviors, this analysis provides a foundation for understanding the cultural effects of digital interaction in India. Ultimately, the findings offer insights into the ways in which social media shapes, and is shaped by, the cultural landscape of the country, illustrating the transformative role of digital communication in modern Indian society.

OBJECTIVES

1. To Analyze Social Media Usage Across Demographics in India:
Identify patterns of social media engagement across age groups, education levels, and professional status to understand generational and social differences in digital behavior**.**
Determine how variations in social media usage across demographics reflect cultural shifts in communication preferences and online engagement**.**

2. To Explore the Relationship Between Education and Social Media Engagement:
Assess the correlation between education level and social media usage to gain insight into how cultural perspectives and information consumption are influenced by education**.**
Identify how different education levels may impact the types of content shared and consumed on platforms, offering insights into knowledge dissemination and cultural values.

3. To Examine Cultural Influence of Social Media Among Different User Types:
Compare the social media behavior of working professionals and students to explore how lifestyle and career priorities influence engagement, cultural trends, and digital interactions.
Assess how these distinctions contribute to broader cultural dialogues, professional representation, and online identity formation.

4. To Investigate the Role of Social Media Presence on Cultural Identity and Influence:
Analyze the effect of follower count and engagement (e.g., posts) on user behavior to understand how social media presence shapes individual and collective identity in Indian society.
Evaluate whether high levels of engagement contribute to cultural influence, trendsetting, or shifts in self-expression norms within Indian social contexts.

5. To Identify Trends in Social Media Usage Over Weekdays and Weekends:
Analyze weekday versus weekend usage to understand cultural patterns in leisure time and digital behavior. Determine how social media fits into daily routines and assess whether certain periods are associated with increased cultural engagement or social expression, indicating shifts in work-life balance and digital leisure trends.

SCOPE

1. **Regional and Demographic Analysis**: Future work could explore cultural variations in social media use across urban and rural areas or different demographic segments within India to capture regional nuances and cultural diversity.

2. **Longitudinal Study of Cultural Evolution**: By tracking data over several years, the project could analyze long-term changes in digital behavior, allowing for a deeper understanding of how social media contributes to evolving cultural norms.

3. **Content-Type Influence**: Extending the analysis to categorize content types (e.g., news, memes, influencer posts) could help reveal which kinds of posts are most influential across demographics and cultural trends.

4. **Sentiment and Emotional Analysis**: Using sentiment analysis on public posts, the study could gauge emotional responses to cultural events, providing insights into how social media reflects collective cultural sentiment.

5. **Cross-Country Comparison**: Comparing India's social media usage with other countries would place these cultural shifts in a global context, highlighting unique or universal social media effects on culture.

6. **Predictive Modeling of Cultural Trends**: Machine learning models like clustering and time-series analysis could predict future engagement patterns, offering foresight into potential cultural shifts driven by social media.

7. **Emerging Technologies**: As new digital formats like virtual reality gain traction, examining their influence on cultural expression could provide insights into the evolving nature of online communication.

8. **Cultural Preservation and Policy Implications**: By studying the effects of social media on traditional practices, policymakers could leverage these insights to promote cultural preservation in the digital age.

9. **Crisis Behavior Analysis**: Analyzing social media use during crises (e.g., natural disasters or social movements) could highlight how collective cultural identity and solidarity are digitally expressed.

10. **Cultural Identity and Psychological Impact**: Exploring social media's role in shaping users' cultural identity and psychological well-being could deepen the understanding of its broader cultural influence.

REVIEW OF LITERATURE

Citations	Methodology	Dataset	Algorithm	Findings
[1] Bali & Desai (2019)	Applied natural language processing (NLP) to classify and analyze fake news propagation patterns in Indian social media. Text data was preprocessed with tokenization, stop-word removal, and stemming. Fake news detection involved sentiment analysis and frequency analysis of misinformation patterns.	Social media posts from Indian platforms, especially focusing on high-traffic fake news stories	NLP techniques like Bag-of-Words (BoW) and TF-IDF vectorization, followed by classification using Decision Trees and Support Vector Machines (SVM)	Found distinct markers in language usage patterns that contribute to fake news dissemination; identified specific keywords and context common to fake news
[2] Chetty & Alathur (2018)	Conducted a systematic review of machine learning methods for detecting and mitigating hate speech in social media. The review categorized hate speech features (e.g., language tone, frequency of offensive terms) and analyzed algorithmic performance in large-scale social media datasets.	Various studies on hate speech in social networks, covering both labeled and unlabeled data	Focused on supervised algorithms such as Logistic Regression, Naïve Bayes, and Deep Neural Networks (DNNs)	Found that DNNs outperformed traditional methods in detecting complex hate speech, especially in multilingual datasets
[3] Dubose (2011)	Descriptive study on the impact of social media, using data analytics to explore engagement patterns. The study uses big data analysis techniques on public social media usage trends to illustrate behavioral shifts over time.	Aggregated public data on social media usage across demographics and time periods	Cluster Analysis and Sentiment Analysis; K-Means clustering and topic modeling via Latent Dirichlet Allocation (LDA)	showed significant shifts in communication preferences and social behavior, with younger demographics engaging most heavily on emerging platforms
[4] Dwivedi et al. (2021)	Methodology included identifying main research themes using big data techniques on text corpora of academic papers.	Academic articles on social media marketing and digital trends	Text mining techniques, such as Named Entity Recognition (NER) and topic modeling; used Word2Vec embeddings for semantic clustering	Suggested predictive analytics and personalization in digital marketing will drive future research.

[5] El Asam & Samara (2016)	Combined psychological and legal frameworks with big data analysis to evaluate cyberbullying patterns. Data preprocessing and feature extraction methods focused on detecting abusive language patterns and frequency of aggression keywords.	Case studies and psychological assessments on cyberbullying incidents	Sentiment analysis with NLP techniques and pattern recognition via Support Vector Machines (SVM)	Identified critical gaps in legal protections for victims; found psychological effects strongly correlated with the frequency and intensity of cyberbullying terms

METHODOLOGY

1. Dataset Acquisition:
   The first step was to identify a dataset relevant to social media usage in India. For this project, we obtained a CSV file containing social media data with attributes like age, education level, professional status, Instagram followers, posts, total social media usage (weekly and weekend), etc. This dataset served as the foundation for our analysis.

2. Data Preprocessing and Cleaning:
   Loading the Data: The dataset was loaded into R using read_csv from the reader package.
   Column Renaming: To make the analysis more accessible, we renamed certain columns to simplify column references in subsequent code blocks.
   Data Filtering: We added an option for filtering by a specific column, making it possible to focus on certain groups, such as "Working Professionals."

3. Data Analysis and Visualisation:
   Histogram for Social Media Usage: We visualized total weekly and weekend social media usage using geom_histogram to understand different engagement patterns.
   Probability Mass Function (PMF): A custom PMF function was created to calculate the probability of each unique usage level, which helps in understanding the likelihood of different usage patterns.
   Violin Plot: We used violin plots to compare total Instagram and Facebook usage, providing insights into distribution and variance.
   Box Plot: Box plots were used to compare distributions of Facebook and Instagram usage, allowing a comparison of median values and spread.
   Scatter Plot Matrix: A scatter plot matrix was generated to explore correlations between multiple variables such as Instagram followers, Facebook usage, and Instagram usage.
   Bar Charts for Educational and Status-Based Analysis: Bar charts were created to analyze social media usage across education levels and user status, helping us understand how these demographic factors influence social media engagement.

4. Interpretation of Results:
   After generating visualizations and exploring different aspects of the data, we drew conclusions based on the observed patterns, such as generational differences in social media use, the correlation between education level and engagement, and distinct patterns in usage between weekends and weekdays.

DATASET

The dataset titled **[6]** "Social Media Usage India" contains information about social media usage patterns across different demographics in India. The characteristics are mentioned below:

1. Overview: Total rows are 1628. Total columns are 26. Data Types are mixed consisting of integer, float and object/string.
2. Key Features: Demographics, Social Media Usage and Additional insights such as Location, Total Social Media Usage, Zone.
3. Data Observations: The dataset has no missing values. Several features like social media usage are recorded as strings, requiring conversion for numerical analysis. The data includes both categorical (e.g., gender, city) and numerical (e.g., age, follower count) variables.

Table 1. Few rows and columns of the dataset

	Age	City	Current Status	Do you own multiple profiles on Instagram?	Gender	Highest Education	Location (City Airport Code)	Phone OS	State	Zone
0	24	Delhi	Working professional	No	Female	Graduation	DEL	iOs	Delhi	Northern
1	39	Delhi	Working professional	No	Female	Post graduation	DEL	iOs	Delhi	Northern
2	22	Mumbai	Working professional	No	Male	Graduation	BOM	Android	Maharashtra	Western
3	26	Bengaluru	Sabbatical	Yes	Female	Graduation	BLR	Android	Karnataka	Southern
4	50	Delhi	Working professional	No	Male	Graduation	DEL	iOs	Delhi	Northern

RESULTS

5 objectives have been achieved with respect to the project:

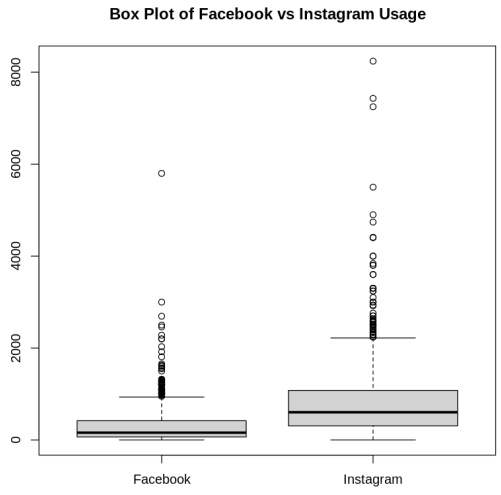
1. **How does social media usage vary among different age groups, and what does this imply about cultural shifts in communication preferences?**
A. Analyzing the total social media usage across age groups can highlight how younger versus older generations engage differently on platforms like Instagram and Facebook, offering insights into generational cultural trends.

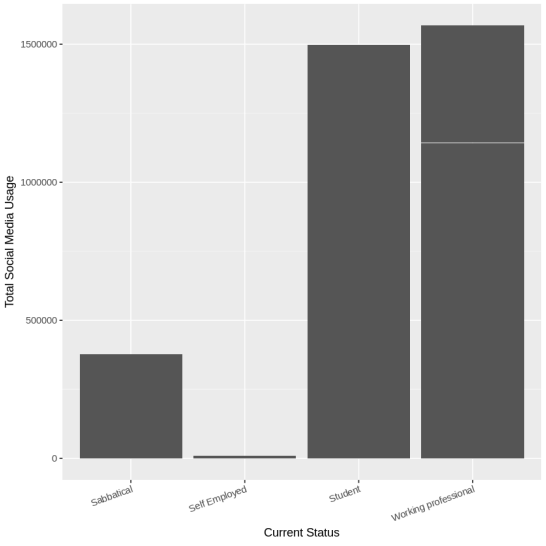
2. **Is there a correlation between education level and social media engagement, and how might this affect cultural perspectives and information consumption?**
By comparing education levels with total social media usage, you can explore whether people with higher education are more or less engaged on social media, possibly reflecting on how informed or culturally diverse perspectives are shared.

3. **How does social media usage differ between working professionals and students, and what cultural implications might these patterns suggest?**
Analyzing social media activity across user status (e.g., working professionals vs. students) could reveal how work-life balance or career priorities influence cultural engagement with online platforms.

4. **What is the impact of follower count and engagement (e.g., posts) on social media on user behavior, and how might this affect cultural influence or identity?**
Investigating if individuals with higher follower counts or post frequency use social media differently may provide insights into how social media affects cultural identity, self-expression, or influence in Indian society.

5. **Are there distinct patterns in weekend vs. weekday social media usage, and what does this suggest about cultural trends in leisure and digital behavior?**
By examining differences in social media usage during weekends versus weekdays, you can explore how social media fits into cultural trends in leisure and digital behavior.





CONCLUSION

The project successfully uses Big Data Analytics within R to uncover insights into how social media is shaping and reflecting cultural shifts across diverse segments of Indian society. Through a systematic analysis of social media usage data, the project offers a nuanced view of online behavior among different age groups, education levels, and professional statuses, shedding light on how digital communication impacts traditional and modern cultural practices in India.

The project's outcomes underscore that social media is not only a platform for communication but also a powerful force shaping and reflecting cultural values in India. Social media usage patterns suggest a shift toward digital integration in daily life, especially among younger generations and educated segments, who may be driving cultural adaptation and influencing norms. This transformation reflects India's ongoing journey towards digital modernization while balancing traditional values.

These findings offer a valuable foundation for future research and policymaking, especially in areas of digital literacy, online behavior, and cultural preservation. By understanding these trends, stakeholders can make informed decisions that promote cultural inclusivity, encourage responsible digital engagement, and foster the positive impacts of social media on Indian society.

REFERENCES

[1] Bali, A., & Desai, P. (2019). Fake News and Social Media: Indian Perspective. Media Watch, 10(3), 737–750.

[2] Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. Aggression and Violent Behavior, 40, 108–118.

[3] Dubose, C. O. (2011). The social media revolution. Radiologic Technology, 83(2), 112–119.

[4] Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, 1–37.

[5] El Asam, A., & Samara, M. (2016). Cyberbullying and the law: A review of psychological and legal challenges. Computers in Human Behavior, 65, 127–141.

[6] Dataset: https://www.kaggle.com/datasets/prachirikhari/social-media-usage-trends-india/data