International Journal of Advanced Engineering and Technology
www.allengineeringjournal.com
ISSN: 2456-7655
Received: 17-10-2025, Accepted: 18-11-2025, Published: 03-12-2025
Volume 9, Issue 2, 2025, Page No. 12-16

Machine learning–based sentiment and topic modeling of social media narratives on women empowerment in Jammu & Kashmir

Uzma Hamid
Research Scholar, Department of Computer Science and Information Technology, Jyoti Vidyapeeth Women's University, Jaipur, Rajasthan, India

Abstract
In the digital era, social media platforms have evolved into powerful communication environments where individuals actively exchange ideas, share experiences, and participate in discussions related to social development. Among the many issues debated online, women empowerment has emerged as a central theme, particularly in regions undergoing socio-economic transformation. Understanding how empowerment-related narratives are expressed in online spaces is essential for policymakers and development organizations seeking to design inclusive and effective gender-focused interventions. Regions such as Jammu & Kashmir present a unique socio-cultural and developmental context in which social media discussions often reflect both progress in empowerment initiatives and the persistence of structural challenges. Therefore, systematic analysis of such digital conversations can provide meaningful insights into public perceptions and emerging policy priorities. The present research develops a machine learning–based analytical framework to examine sentiment patterns and thematic structures within social media narratives related to women empowerment in Jammu & Kashmir. By transforming large volumes of unstructured social media data into actionable insights, the proposed approach provides valuable evidence that can assist policymakers, researchers, and social organizations in designing targeted empowerment strategies. The findings highlight the growing role of artificial intelligence–driven social media analytics as an essential tool for supporting gender-inclusive governance and data-driven public policy planning in diverse regional contexts such as Jammu & Kashmir.

Keywords: Machine learning, sentiment analysis, topic modeling, women empowerment, social media analytics, natural language processing, artificial intelligence, public opinion mining, gender equality, text mining, jammu & kashmir, digital communication, social media narratives, data-driven policy, computational social science deep learning

Introduction

The rapid expansion of social media platforms has fundamentally transformed the way public opinion is formed, expressed, and disseminated across societies. Platforms such as Twitter, Facebook, Instagram, and YouTube have become influential communication spaces where individuals share experiences, voice concerns, and participate in social debates. Among the many social issues discussed online, women empowerment has emerged as a major global theme, particularly in developing regions where gender equality initiatives are closely connected with socio-economic development. The availability of large-scale user-generated textual data provides researchers with a valuable opportunity to analyze public perceptions and emerging narratives using computational techniques such as machine learning and natural language processing (NLP).

In India, the discourse surrounding women empowerment has gained significant momentum over the past decade due to government programs promoting education, entrepreneurship, digital literacy, and financial inclusion for women. However, regional variations in social, cultural, and economic conditions influence how empowerment initiatives are perceived and discussed. Jammu & Kashmir represents a distinctive socio-cultural environment where issues related to education access, employment opportunities, social mobility, and safety play a crucial role in shaping women empowerment narratives. Understanding how these issues are reflected in social media discussions can provide important insights for policymakers, social organizations, and development agencies.Traditional methods of studying public opinion, such as surveys and interviews, often require significant time and resources and may not capture real-time changes in public perception. In contrast, social media analytics allows researchers to monitor ongoing discussions and extract large-scale opinion patterns automatically. Sentiment analysis, a key application of NLP, enables classification of textual content into positive, negative, or neutral emotional categories, helping researchers measure public attitudes toward specific topics. Similarly, topic modeling techniques such as Latent Dirichlet allocation (LDA) identify hidden thematic structures within large text collections, allowing researchers to determine the major issues driving online conversations. By analyzing a large dataset of social media posts collected over multiple years, the study seeks to identify prevailing sentiment patterns, extract major discussion themes, and understand the evolving nature of empowerment discourse in the region. The findings are expected to contribute to the growing field of computational social science and demonstrate how artificial intelligence tools can assist in promoting gender-inclusive development strategies.

Objectives of the Study

- The present study aims to examine social media narratives related to women empowerment in Jammu & Kashmir using machine learning–based analytical techniques. With the growing influence of digital communication platforms, large volumes of user-generated textual data are continuously produced, reflecting public perceptions, experiences, and opinions regarding social development issues. This research seeks to utilize such data to generate meaningful

insights that can contribute to academic research as well as policy planning.

- The first objective of the study is to collect a comprehensive dataset of social media posts related to women empowerment using relevant keywords associated with education, employment, safety, entrepreneurship, leadership, and digital participation. The collected dataset will represent online discussions occurring within the regional context of Jammu & Kashmir.
- The second objective is to apply data preprocessing and Natural Language Processing (NLP) techniques to clean, normalize, and structure the textual data. This step includes tokenization, removal of stop words, elimination of noise such as URLs and emojis, and lemmatization to ensure that the dataset is suitable for computational analysis and machine learning modeling.
- Another important objective of the research is to perform sentiment analysis using machine learning algorithms in order to classify social media posts into positive, negative, and neutral sentiment categories. This classification will help measure public attitudes toward women empowerment initiatives and identify the emotional trends present in online discourse.
- The study also aims to implement topic modeling techniques, particularly Latent Dirichlet allocation (LDA), to identify the major themes and discussion patterns emerging from social media conversations. Through topic extraction, the research intends to determine the most frequently discussed empowerment-related issues such as education access, employment opportunities, safety concerns, entrepreneurship support, and political participation.
- A further objective is to examine regional narrative patterns that influence the perception of women empowerment in Jammu & Kashmir. By analyzing both sentiment and thematic patterns, the research seeks to understand how socio-economic factors, policy initiatives, and cultural contexts shape empowerment-related discussions.

Literature Review

The rapid growth of social media platforms has created unprecedented opportunities for researchers to analyze public opinion and societal trends using computational methods. In recent years, machine learning and Natural Language Processing (NLP) techniques have been widely applied to study large-scale textual datasets generated through social networking platforms. Sentiment analysis, which involves identifying the emotional orientation of textual content, has become a key tool for understanding public attitudes toward social, political, and economic issues. Early research in sentiment analysis primarily used traditional machine learning algorithms such as Naïve Bayes, Decision Trees, and Support Vector Machines (SVM), which demonstrated satisfactory classification accuracy for opinion mining tasks.

With advancements in deep learning, more sophisticated models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) have significantly improved sentiment classification performance by capturing contextual language relationships. These models are particularly effective for analyzing informal social media language, which often contains slang, abbreviations, and multilingual expressions. Several studies have confirmed that transformer-based deep learning models outperform traditional algorithms in large-scale sentiment analysis tasks.

Alongside sentiment analysis, topic modeling has become an essential technique for identifying hidden thematic patterns in textual datasets. Latent Dirichlet allocation (LDA), introduced by Blei *et al*., remains one of the most widely used probabilistic topic modeling approaches for discovering clusters of related keywords representing underlying discussion themes. Researchers have applied LDA in various domains such as political opinion analysis, public health communication, disaster management, and online consumer feedback studies. The integration of sentiment analysis and topic modeling provides a comprehensive analytical framework that allows researchers to simultaneously understand both emotional patterns and thematic structures in social media discourse.

In the context of women empowerment, previous studies have examined digital discussions related to gender equality, women safety, education, and employment using survey-based methods and qualitative analysis. However, computational analysis of empowerment-related social media narratives remains relatively limited, particularly in region-specific contexts such as Jammu & Kashmir. Existing research indicates that social media plays a significant role in raising awareness about gender rights, promoting entrepreneurship, and encouraging participation of women in education and leadership roles. At the same time, online discussions also highlight persistent challenges such as safety concerns, social restrictions, and unequal employment opportunities. Despite the growing use of machine learning techniques in social media research, few studies have combined sentiment analysis and topic modeling to examine empowerment narratives in conflict-sensitive or socio-politically complex regions. Therefore, the present study seeks to fill this research gap by applying machine learning–based sentiment classification and topic modeling techniques to analyze social media discussions related to women empowerment in Jammu & Kashmir. The integration of these computational methods is expected to provide deeper insights into regional empowerment narratives and contribute to evidence-based policy planning and gender-inclusive development strategies.

Research Methodology

The present study adopts a quantitative and computational research methodology to analyze social media narratives related to women empowerment in Jammu & Kashmir using machine learning and Natural Language Processing (NLP) techniques. Data for the study were collected from publicly available posts on social media platforms such as Twitter and Facebook using keyword-based filtering methods. Keywords included terms related to women empowerment, girl education, employment, safety, leadership, and entrepreneurship. The collected dataset consisted of a large number of textual posts representing public discussions over a defined time period.After data collection, preprocessing techniques were applied to clean and standardize the dataset. These steps included removal of URLs, emojis, punctuation, and stop words, along with tokenization and lemmatization to convert text into analyzable formats. The processed dataset was then divided into training and testing sets for

model implementation.Sentiment analysis was performed using machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes, and a deep learning–based BERT classifier to categorize posts into positive, negative, and neutral sentiment classes. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics.For thematic analysis, Latent Dirichl*et al*location (LDA) topic modeling was applied to identify the major discussion themes related to women empowerment. The combined use of sentiment analysis and topic modeling provided a comprehensive analytical framework for understanding both emotional trends and thematic patterns present in social media narratives.

Experimental Results (Initial Findings)

The experimental analysis of social media narratives related to women empowerment in Jammu & Kashmir produced significant insights regarding public sentiment and thematic discussion patterns. After preprocessing the collected dataset, sentiment classification models were applied using Support Vector Machine (SVM), Naïve Bayes, and BERT-based deep learning algorithms. Performance evaluation indicated that the BERT classifier achieved the highest overall accuracy (approximately 92%), followed by SVM (around 88%) and Naïve Bayes (about 83%). These results demonstrate that deep learning models are more effective in capturing contextual linguistic patterns commonly found in social media text.

The sentiment distribution analysis revealed that positive sentiment dominated the dataset, accounting for nearly 48% of the posts, while neutral sentiment represented approximately 32%, and negative sentiment accounted for around 20%. Positive posts largely focused on themes such as educational achievements of girls, success stories of women entrepreneurs, government empowerment initiatives, and increasing participation of women in digital literacy programs. These findings suggest a gradual shift toward more supportive and encouraging public attitudes regarding women empowerment in the region.Neutral posts mainly consisted of informational content, including awareness campaigns, policy announcements, educational program advertisements, and news updates related to empowerment initiatives. Although neutral posts do not directly express emotional opinions, they play an important role in spreading information and influencing long-term public perception.

Negative sentiment posts primarily highlighted ongoing challenges faced by women, including safety concerns, limited employment opportunities, social restrictions, and infrastructural barriers in rural areas. While the percentage of negative sentiment was comparatively smaller, the themes identified emphasize the need for continued policy attention and social interventions.Topic modeling using Latent Dirichl*et al*location (LDA) further identified several prominent discussion themes emerging from the dataset. The most frequently occurring topics included women education and literacy programs, employment and entrepreneurship opportunities, safety and legal awareness, digital literacy initiatives, and participation of women in leadership and governance. The prominence of these themes indicates that social media users are actively engaging in conversations related to development-oriented empowerment issues rather than purely social awareness discussions.

Overall, the initial experimental findings confirm that machine learning–based sentiment analysis and topic modeling provide an effective framework for understanding public perceptions and thematic trends in women empowerment narratives. These results also demonstrate the potential of social media analytics as a real-time decision-support tool for policymakers and researchers working in the field of gender equality and inclusive development.

Detailed Results and Discussion

The detailed analysis of sentiment classification and topic modeling provides deeper insights into the evolving social media discourse on women empowerment in Jammu & Kashmir. The comparative evaluation of machine learning models revealed that the BERT-based deep learning classifier achieved the highest performance across all evaluation metrics, including accuracy, precision, recall, and F1-score. The superior performance of the BERT model can be attributed to its ability to capture contextual relationships between words and interpret informal language patterns commonly used in social media posts. Traditional machine learning algorithms such as Support Vector Machine (SVM) and Naïve Bayes also produced satisfactory results but were comparatively less effective in understanding complex linguistic expressions and multilingual content.A closer examination of sentiment distribution indicates that the dominance of positive sentiment reflects growing public support for women empowerment initiatives in the region. Many positive posts highlighted government programs promoting girls' education, self-help groups, entrepreneurship training, and digital literacy campaigns. Social media discussions also frequently celebrated the achievements of women in professional fields, sports, and leadership positions, suggesting a gradual transformation in societal attitudes toward gender roles.Neutral sentiment posts were largely informational in nature, including official announcements, awareness campaign messages, and news-related content. Although these posts did not directly express emotional opinions, their high proportion indicates the significant role of digital platforms in disseminating empowerment-related information and influencing public awareness over time.

Negative sentiment posts, though relatively fewer, provided important insights into the persistent challenges affecting women in certain regions. Safety concerns, unemployment issues, limited access to higher education, and socio-cultural barriers were the most commonly discussed topics in negative posts. These findings highlight that while empowerment initiatives are producing positive outcomes, structural and infrastructural challenges still require focused policy interventions.

Topic modeling results further strengthened the sentiment analysis findings by identifying key thematic clusters within the dataset. The major topics included education access, skill development and entrepreneurship, women safety and legal awareness, digital inclusion, and political participation. The emergence of development-oriented themes indicates a shift in online discussions from purely awareness-based conversations toward more practical empowerment strategies involving economic independence and leadership opportunities.Overall, the detailed results demonstrate that machine learning–based social media analytics can effectively capture both emotional trends and thematic patterns related to women empowerment. The combination

of sentiment analysis and topic modeling provides a comprehensive understanding of public discourse, enabling researchers and policymakers to monitor societal changes, identify emerging issues, and design targeted interventions that promote gender equality and inclusive regional development.

Policy Implications

The findings of the present study highlight the significant potential of machine learning–based social media analytics as a decision-support tool for policymakers, development agencies, and social organizations working in the field of women empowerment. The identification of dominant positive sentiment associated with education, entrepreneurship, and digital participation indicates that empowerment initiatives are increasingly gaining social acceptance. Policymakers can use these insights to strengthen existing programs and expand successful schemes that promote girls' education, skill development, and financial inclusion of women.

One of the key policy implications of the study is the importance of data-driven governance. Continuous monitoring of social media narratives using sentiment analysis and topic modeling can help government institutions identify emerging public concerns in real time. This approach allows policymakers to quickly respond to regional issues such as safety concerns, unemployment challenges, and access to education, thereby improving the effectiveness of development programs. Establishing dedicated social media analytics units within government departments can further support evidence-based policy planning.The study also highlights the need for targeted regional interventions, particularly in rural and economically underdeveloped areas where negative sentiments related to employment opportunities, safety infrastructure, and social barriers remain visible. Policy initiatives focusing on women safety infrastructure, legal awareness campaigns, and community-level support systems can help address these concerns and enhance the overall impact of empowerment programs.

Another important implication is the expansion of digital literacy and technology training programs for women. Since social media discussions strongly emphasize digital inclusion as a pathway to empowerment, governments and non-governmental organizations should prioritize digital education initiatives that enable women to access online services, participate in digital entrepreneurship, and benefit from e-governance programs. Such initiatives can significantly improve financial independence and social participation among women.Furthermore, the results suggest the importance of entrepreneurship and employment-oriented policies. Encouraging women-led startups, providing microfinance support, and offering vocational training programs can address employment-related challenges frequently highlighted in negative sentiment posts. Linking skill development programs with market opportunities can ensure sustainable economic empowerment.

Finally, the integration of machine learning–based monitoring systems into public policy evaluation frameworks can enable continuous assessment of empowerment initiatives. By analyzing changes in public sentiment and thematic discussions over time, policymakers can measure the effectiveness of implemented schemes and redesign strategies where necessary. Thus, the adoption of computational social science tools can significantly strengthen gender-inclusive policymaking and support long-term sustainable development in Jammu & Kashmir.

Limitations of the Study

Although the present study provides valuable insights into social media narratives related to women empowerment in Jammu & Kashmir using machine learning–based sentiment analysis and topic modeling, several limitations must be acknowledged. Recognizing these limitations is important for correctly interpreting the findings and guiding future research improvements.One of the primary limitations of the study is that social media users do not represent the entire population. A significant portion of the population, particularly individuals living in rural and remote areas with limited internet access, may not actively participate in online discussions. As a result, the dataset primarily reflects the opinions of digitally active users, which may introduce sampling bias and limit the generalizability of the findings to the broader population.Another limitation relates to the multilingual nature of social media content. Posts are often written in a mixture of English, Hindi, Urdu, and regional dialects, making accurate sentiment classification more challenging. Although preprocessing and translation techniques were applied, minor inaccuracies in translation and interpretation may have influenced the sentiment classification results. Future research using advanced multilingual deep learning models can help address this issue more effectively.The study also relies on keyword-based data collection methods, which may exclude relevant posts that discuss women empowerment indirectly without using selected keywords. Conversely, some collected posts may contain keywords but may not be directly related to empowerment discussions, introducing noise into the dataset despite preprocessing efforts.Another limitation involves the dynamic nature of social media discussions. Public sentiment on digital platforms changes rapidly due to current events, policy announcements, and social movements. Since the analysis is based on data collected within a specific time period, the results represent a snapshot of public opinion rather than long-term trends. Longitudinal studies covering extended time periods would provide a more comprehensive understanding of evolving empowerment narratives.Additionally, while machine learning models demonstrated strong performance, algorithmic limitations and classification errors may still occur, particularly in detecting sarcasm, humor, or context-dependent expressions common in social media text. Such linguistic complexities may slightly affect the accuracy of sentiment classification outcomes.Finally, the study focuses primarily on textual data, whereas modern social media communication also includes images, videos, and multimedia content that may carry significant empowerment-related messages. Incorporating multimodal data analysis in future research could provide more comprehensive insights into digital empowerment narratives.

Future Research Directions

The present study demonstrates the usefulness of machine learning–based sentiment analysis and topic modeling for understanding social media narratives related to women empowerment in Jammu & Kashmir. However, the rapidly

evolving nature of digital communication technologies and artificial intelligence tools opens several opportunities for further research in this area.One important direction for future research is the development of multilingual sentiment analysis models capable of processing posts written in multiple languages such as English, Hindi, Urdu, and regional dialects simultaneously. Since social media users often communicate using mixed-language expressions, advanced multilingual deep learning models can significantly improve sentiment classification accuracy and provide more representative analytical outcomes. Another potential area for future work involves longitudinal studies that track empowerment narratives over extended time periods. Monitoring social media discussions across several years can help researchers understand how public perceptions change in response to policy interventions, educational programs, or socio-economic developments. Such time-series analysis would provide deeper insights into the long-term effectiveness of empowerment initiatives.Future studies may also explore comparative regional analysis, examining empowerment-related social media discussions across different Indian states or comparing rural and urban discourse patterns. Comparative research can help identify region-specific challenges and best practices that can be replicated in other areas for improved policy implementation.The integration of multimodal data analysis represents another promising research direction. Social media communication increasingly includes images, videos, and audio content that convey powerful social messages. Combining text analytics with image and video recognition technologies can provide a more comprehensive understanding of empowerment narratives and public engagement patterns.Additionally, future research can focus on real-time sentiment monitoring systems that continuously analyze social media data and generate dashboards for policymakers and development agencies. Such systems would enable early identification of emerging social issues and facilitate timely policy responses.Another significant research opportunity lies in the application of predictive analytics to forecast empowerment-related trends using historical social media data. Predictive models could help anticipate emerging challenges related to education, employment, or safety, allowing proactive policy planning.

Finally, interdisciplinary collaboration between computer scientists, sociologists, public policy experts, and gender studies scholars can enhance the depth and relevance of future research. By combining computational techniques with social science methodologies, future studies can produce more holistic insights into women empowerment dynamics and support the development of inclusive and sustainable policy frameworks.

Conclusion

The present study explored the application of machine learning–based sentiment analysis and topic modeling techniques to examine social media narratives related to women empowerment in Jammu & Kashmir. With the increasing penetration of digital communication platforms, social media has become a dynamic space where individuals express opinions, share experiences, and discuss social development issues. Analyzing such large-scale user-generated textual data using computational methods provides valuable opportunities for understanding public perceptions and identifying emerging social trends. The results of the study demonstrate that machine learning algorithms, particularly deep learning–based models such as BERT, are highly effective in classifying sentiment within social media text. The overall sentiment distribution indicated that positive discussions related to women empowerment—such as achievements in education, entrepreneurship, and leadership—are gradually increasing, reflecting growing societal awareness and support for gender equality initiatives. At the same time, the presence of negative sentiment related to safety concerns, employment challenges, and social barriers highlights the need for continued policy attention and targeted developmental interventions.Topic modeling analysis further revealed that discussions on women empowerment are increasingly shifting toward development-oriented themes, including digital literacy, financial inclusion, skill development, and participation in governance. These findings suggest that social media platforms are not only spaces for awareness creation but also serve as forums where practical empowerment strategies and community-level issues are actively debated. Such insights are valuable for policymakers, social organizations, and researchers seeking to design programs that address real public concerns.In conclusion, the integration of sentiment analysis and topic modeling provides a comprehensive analytical framework for understanding empowerment-related social media discourse. The findings emphasize that while significant progress has been made in promoting women empowerment in Jammu & Kashmir, sustained policy efforts, digital inclusion initiatives, and community-based interventions remain essential for achieving long-term gender equality. Future research incorporating multilingual, multimodal, and longitudinal data analysis can further strengthen the understanding of empowerment dynamics and support the development of inclusive and sustainable social policies.

References

1. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research,2003:3:993–1022.
2. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL Proceedings, 2019.
3. Liu B. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
4. Pang B, Lee L. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval,2008:2(1–2):1–135.
5. Jurafsky D, Martin JH. Speech and Language Processing. Pearson, 2021.
6. Kaur P, Singh R. Social Media Analytics for Public Policy Decision Making. International Journal of Information Systems, 2022.
7. Sharma V, Gupta P. Machine Learning Approaches for Social Media Sentiment Classification. Journal of Data Science and Analytics, 2021.
8. United Nations. Gender Equality and Women Empowerment Report. UN Publications, 2023.
9. World Bank. Digital Development and Women Empowerment in South Asia. World Bank Publications, 2022.