Sentiment Analysis of Tweets on Afghan Women's Rights Using Naive Bayes Classifier: A Data Mining Approach to Understanding Public Discourse

Thosporn Sangsawang[1,*], Liu Yang[2]

[1]Educational Technology and Communications Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand

[2]Vocational Education Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand

ABSTRACT

Social media platforms have become critical arenas for public discourse on global human rights issues, providing real-time insight into public opinion and emotional responses. This study examines Twitter conversations surrounding Afghan women's rights from April 2023 to January 2024, focusing on the digital reflection of international concern. Using a dataset of 4,845 cleaned English-language tweets, we performed sentiment analysis employing the VADER lexicon for initial sentiment labeling and a Multinomial Naive Bayes classifier trained on TF-IDF features for automated sentiment classification. The results reveal a predominance of negative sentiment (47.4%) compared to positive (38.3%) and neutral (14.3%) sentiments, indicating widespread frustration and alarm regarding the restrictions and violations faced by Afghan women. Exploratory data analysis highlighted temporal trends in tweet volume and engagement, with significant peaks correlating to key political events and policy announcements. The model achieved an overall accuracy of 67.5% in classifying sentiment, with particularly strong performance in detecting negative and positive tweets, while neutral sentiments were more challenging to classify accurately. Feature importance analysis identified critical terms that influenced sentiment classification, revealing a linguistic pattern reflective of advocacy, concern, and hope within the discourse. Temporal analysis of sentiment proportions demonstrated fluctuations aligning with real-world developments, underscoring the dynamic nature of online public opinion. This research contributes to understanding the role of social media in amplifying human rights concerns, especially in politically unstable regions, and demonstrates the utility of sentiment analysis for monitoring global digital activism. The findings offer valuable insights for policymakers, activists, and scholars interested in the intersection of technology, public opinion, and human rights advocacy. Future research is encouraged to incorporate multilingual data, multiple social media platforms, and analyze sentiment shifts in response to international interventions to provide a more comprehensive picture of digital society engagement on this critical issue.

Keywords Afghan Women's Rights, Sentiment Analysis, Social Media, Naive Bayes Classifier, Digital Activism

Submitted 26 May 2025
Accepted 1 June 2025
Published 2 June 2025

*Corresponding author
Thosporn Sangsawang,
sthosporn@rmutt.ac.th

Additional Information and Declarations can be found on page 179

Introduction

In recent years, social media has emerged as a powerful platform for expressing opinions, discussing social issues, and mobilizing change. The transformative nature of social media is evident in its ability to facilitate rapid information dissemination and foster community engagement. This capability has been particularly instrumental in political activism, where social media serves as a

How to cite this article: T.Sangsawang and L.Yang, "Sentiment Analysis of Tweets on Afghan Women's Rights Using Naive Bayes Classifier: A Data Mining Approach to Understanding Public Discourse," *J. Digit. Soc.*, vol. 1, no. 2, pp. 168-182, 2025.

catalyst for organizing protests, raising awareness, and mobilizing support for various causes. Keith emphasizes that the effectiveness of social media in driving political change is contingent on the strategic alignment of online campaigns with offline actions, highlighting the importance of synergizing digital advocacy with physical participation [1]. This synergy underscores the role of social media not merely as a communication tool but as a vital component of contemporary activism.

Moreover, social media platforms have significantly influenced public perceptions and behaviors, particularly among adolescents. Research indicates that high social media usage is associated with increased high-risk behaviors, including substance use and early sexual activity [2], [3]. Adolescents often seek social validation through their online interactions, which can lead to engagement in risky behaviors as they attempt to conform to perceived peer norms. This phenomenon illustrates the dual-edged nature of social media, where it can foster community and connection while simultaneously exposing users to negative influences and pressures.

Furthermore, the psychological effects of social media usage are profound, particularly among younger populations. Studies have identified a correlation between excessive social media use and mental health issues, including anxiety and depression [4], [5]. The addictive nature of social media can lead to detrimental outcomes, emphasizing the need for social media literacy programs aimed at empowering users to navigate these platforms responsibly [6]. By fostering awareness of the potential risks associated with social media, individuals can better manage their online interactions and mitigate negative impacts on their well-being.

Social media platforms, particularly Twitter, have become essential tools for real-time engagement on global topics, including human rights. The immediacy and accessibility of Twitter allow users to share information, mobilize support, and raise awareness about human rights issues almost instantaneously. This capability has been recognized as a significant factor in enhancing civic engagement and political participation, as social media facilitates the rapid dissemination of information and fosters community discussions around pressing social issues [7].

Moreover, the role of social media in addressing human rights concerns is underscored by the responsibilities that platforms like Twitter hold in moderating content related to gender-based violence and other human rights violations. Research indicates that social media platforms must be designed to respect and promote human rights, which includes implementing policies that address online harassment and violence [8]. This responsibility is critical, as the digital infrastructure increasingly shapes public discourse and influences societal norms regarding human rights.

The impact of social media on public health and human rights advocacy has also been highlighted during global crises, such as the COVID-19 pandemic. Social media has served as a vital channel for disseminating information about public health measures and the human rights implications of such measures, allowing marginalized voices to be heard when traditional media may overlook them [9]. The ability of social media to amplify these voices is crucial in contexts where individuals may lack access to other forms of communication or advocacy platforms.

Furthermore, the engagement on platforms like Twitter is not limited to information sharing; it also encompasses the mobilization of grassroots movements. Social media has been instrumental in organizing protests and campaigns, as evidenced by movements such as #MeToo and Black Lives Matter, which gained significant traction through Twitter [10]. These movements illustrate how social media can serve as a catalyst for social change, enabling users to connect, share experiences, and advocate for justice and equality on a global scale.

Afghan women's rights have remained a pressing concern in the international human rights agenda. Over the past few years, global conversations have increasingly centered on the struggles faced by women in Afghanistan, particularly in accessing education, employment, and the freedom to participate fully in public life. These fundamental rights, often taken for granted in many parts of the world, are persistently under threat in Afghanistan due to longstanding sociopolitical and cultural restrictions.

The reemergence of the Taliban regime in 2021 has further exacerbated the already fragile status of women’s rights in the country. Restrictions imposed on girls' education, employment bans for women in certain sectors, and limitations on their mobility have sparked international outcry. In the absence of free press and civic space within Afghanistan, social media platforms—especially Twitter—have become vital spaces for sharing personal testimonies, news coverage, and calls for international solidarity.

Social media, therefore, is not only a reflection of public sentiment but also a tool for mobilizing global support. The digital discourse surrounding Afghan women's rights reveals real-time emotional responses and collective concerns expressed by users worldwide. In this context, analyzing social media posts becomes a powerful means of capturing public opinion and tracking the dynamics of global awareness over time.

The objective of this study is to examine public sentiment related to Afghan women’s rights by conducting a sentiment analysis on tweets collected from April 2023 to January 2024. By processing and analyzing the text content of these tweets using natural language processing techniques and machine learning algorithms, this research seeks to classify and understand the emotional tone (positive, negative, neutral) surrounding the topic during this critical period.

Such an analysis is significant because it provides empirical insight into how the international digital community perceives the status of Afghan women. The findings may reveal peaks in concern or support that coincide with key events or policies and highlight the issues that resonate most with the global public. This can inform the design of advocacy campaigns, media strategies, and humanitarian interventions.

Literature Review

Social Media and Public Opinion

Social media platforms, particularly Twitter, have become essential tools for public discourse, especially in the context of political and social issues. The immediacy and accessibility of Twitter allow users to engage in real-time discussions, share opinions, and mobilize support for various causes. This dynamic environment has made Twitter a significant platform for political engagement and activism, as it enables users to express their views and organize collective actions swiftly [11], [12]. Research indicates that Twitter serves as a valuable resource for understanding public opinion, particularly during political events. For instance, the analysis of digital trace data from Twitter can reveal insights into how political actors are perceived and how public sentiment evolves over time [11]. Moreover, Twitter's ability to foster opinion leadership plays a crucial role in civic participation. Influential users can drive discussions on social issues, encouraging broader public engagement and participation in the discourse [13]. This phenomenon illustrates how social media can democratize information dissemination and empower individuals to contribute to public debates, transcending traditional barriers related to socioeconomic status or access to information.

Analyzing Twitter data has proven to be a valuable method for gaining insights into public sentiment and social issues. Numerous studies have demonstrated the effectiveness of sentiment analysis on Twitter, particularly in the context of significant events such as the COVID-19 pandemic. For instance, Elyashar et al. conducted a text analysis of Twitter discourses to assess the mental state of healthcare professionals during the pandemic, revealing critical insights into their experiences and concerns [14]. This highlights how Twitter can serve as a rich source of qualitative data that reflects the sentiments of specific populations in response to societal challenges. Furthermore, Boon-Itt and Skunkan's research on public perception of the COVID-19 pandemic utilized sentiment analysis and topic modeling to analyze Twitter data, emphasizing the platform's utility in understanding collective attitudes towards health crises [15]. Their findings illustrate how Twitter can capture the nuances of public sentiment, providing a real-time barometer of societal reactions to unfolding events.

Sentiment Analysis

Sentiment analysis in text mining is a crucial technique that involves classifying text data into positive, negative, or neutral categories based on the underlying emotions or opinions expressed within the text. This approach has gained significant traction in various fields, particularly in analyzing social media content, where user-generated text often reflects public sentiment on a wide range of topics, including political events, consumer products, and social issues [16], [17]. Similarly, [18] explored the methodologies used in sentiment analysis, noting the development of various algorithms designed to identify subjective opinions expressed in online texts.

The application of sentiment analysis extends beyond mere classification; it also involves understanding the context and nuances of the text being analyzed. Cahyani & Patasik compared different models for emotion text classification, illustrating how various techniques can yield different insights depending on the approach taken [19]. Additionally, Pahwa et al. discussed the significance of pre-processing techniques in sentiment analysis, which are essential for improving the accuracy of classification tasks [20].

Different machine learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), and Deep Learning models, have been applied to sentiment analysis with varying degrees of success. Each of these algorithms has its strengths and weaknesses, making them suitable for different types of sentiment analysis tasks.

Naive Bayes is one of the most commonly used algorithms for sentiment analysis due to its simplicity and effectiveness. It operates on the principle of Bayes' theorem and assumes that the features are independent given the class label. Studies have shown that Naive Bayes can achieve high accuracy in sentiment classification tasks. For example, Saputro et al. utilized Naive Bayes in their analysis of public sentiment during the 2018 governor's election, highlighting its fast processing time and ease of implementation [21]. Similarly, Mustolih applied Naive Bayes for sentiment analysis in automotive reviews, confirming its effectiveness in handling textual data [22]. However, while Naive Bayes is efficient, it may struggle with complex datasets where feature independence is not a valid assumption.

Support Vector Machines (SVM) have also been widely used in sentiment analysis, particularly due to their ability to handle high-dimensional data and their effectiveness in binary classification tasks. Lutfi et al. compared SVM with Naive Bayes in the context of sales reviews, finding that SVM often outperformed Naive Bayes in terms of accuracy [23]. Rezki et al. further demonstrated the robustness of SVM in sentiment analysis, achieving high accuracy rates when applied to airline reviews [24]. The SVM's strength lies in its capacity to find the optimal hyperplane that separates different classes, making it particularly effective for sentiment classification tasks where the data is not linearly separable.

Deep Learning models have emerged as powerful tools for sentiment analysis, leveraging neural networks to capture complex patterns in data. These models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have shown remarkable performance in various sentiment analysis tasks. For instance, Chakravarthy discussed the use of hybrid architectures that combine traditional machine learning approaches with deep learning techniques, indicating that deep learning can significantly enhance sentiment classification accuracy [25]. The ability of deep learning models to learn hierarchical representations of data allows them to outperform traditional algorithms, especially in large datasets with intricate patterns.

Afghan Women's Rights in Social Media

Studies have explored how social media has been utilized to raise awareness about the human rights of Afghan women, particularly following the Taliban's takeover in 2021. The resurgence of the Taliban has led to severe restrictions on women's rights, prompting activists and organizations to leverage social media platforms to highlight these issues and mobilize support for Afghan women [26].

One significant study by Lalyzada emphasizes the role of media and social networks in promoting the position of Afghan women, suggesting that social media has become a vital tool for advocacy and awareness-raising in the face of oppressive regimes [27]. The study argues that social media can help Afghan women articulate their struggles and aspirations, thereby fostering a sense of community and solidarity among those advocating for gender equality. This aligns with findings from Khalili, who notes that social media serves as a powerful instrument for overcoming traditional barriers and promoting women's rights in Afghanistan [28].

Becvar's research further supports the notion that social media can influence attitudes toward gender equality in Taliban-controlled Afghanistan. The study indicates that public discourse on social media can shape perceptions and foster a more supportive environment for women's rights advocacy [29]. This reflects a broader trend where social media acts as a platform for human rights advocacy, allowing individuals and organizations to connect, share information, and mobilize resources effectively.

Gaps in Existing Research

While there is a growing body of literature on sentiment analysis and public opinion, there is a notable lack of studies specifically focusing on the intersection of social media discourse, Afghan women's rights, and sentiment analysis. This gap is particularly significant given the recent political changes in Afghanistan and the urgent need to understand public sentiment regarding the rights of Afghan women.

Nguyen et al. highlight the limitations of sentiment analysis in capturing nuanced social issues, suggesting that while machine learning techniques can categorize tweets by sentiment, they may not fully address the complexities of the topics discussed [30]. This observation is relevant to the study of Afghan women's rights, where the discourse is often intertwined with cultural, political, and social dynamics that require a more sophisticated analytical approach.

Manganaro and Alozie discuss the broader context of women's rights in Afghanistan, noting that legal guarantees do not always translate into actual empowerment [31]. This highlights the need for sentiment analysis to explore public perceptions of these rights and the effectiveness of advocacy efforts on social media. Understanding how the public feels about Afghan women's rights can inform strategies for raising awareness and mobilizing support.

Afrouz et al. provide qualitative insights into the barriers Afghan women face, but again, the integration of sentiment analysis could enhance these findings by quantifying public sentiment towards these barriers and the proposed solutions [32]. This quantitative approach could complement qualitative narratives, offering a more comprehensive view of the discourse surrounding Afghan women's rights.

Moreover, while some studies have employed sentiment analysis in other contexts, such as maternal health during the COVID-19 pandemic [33], similar methodologies have not been applied to the specific context of Afghan women's rights. This presents an opportunity for future research to fill the existing gap by systematically analyzing tweets related to Afghan women's rights, thereby providing valuable insights into public sentiment and discourse.

Method

Data Collection and Setup

The dataset used in this study, titled "Tweets about Afghan Women," was obtained in CSV format, containing tweets posted from April 2023 to January 2024. Data loading was conducted using the Python library pandas, primarily with UTF-8 encoding. In cases where encoding errors occurred, an alternative ISO-8859-1 encoding was attempted to ensure robustness. To facilitate reproducibility and efficient workflow, directories were programmatically created to store intermediate datasets, machine learning models, vectorizers, evaluation metrics, and visualization outputs. Essential libraries such as NLTK (for natural language processing), scikit-learn (for machine learning), and visualization libraries including matplotlib, seaborn, and wordcloud were imported and configured. Plotting parameters were standardized, with default figure sizes set to 12 by 6 inches and seaborn's "whitegrid" style applied for clarity in visual outputs.

Exploratory Data Analysis (EDA)

Initial data inspection involved assessing the dataset's shape and completeness using the info() method, with verbosity enabled to include detailed data types and missing value counts. Date-time standardization was performed on the 'date' column using pandas.to_datetime() with error coercion enabled (errors='coerce'), dropping any rows that could not be parsed, which accounted for less than 1% of the data. The dataset was filtered to retain only tweets labeled as English (lang == 'en') to ensure compatibility with the English-centric sentiment tools and stopword lists. Temporal analysis included resampling tweets by month (resample('M')) to visualize tweet volume over time, revealing periodic spikes in user engagement. Distribution of engagement metrics, specifically 'likeCount' and 'retweetCount,' were visualized using log-transformed histograms (log1p) to account for highly skewed data distributions. Top tweet sources, such as "Twitter Web App" and "Twitter for iPhone," were analyzed by counting occurrences within the 'sourceLabel' column, and the top 10 sources were presented via horizontal bar plots.

Text Preprocessing

Preprocessing of tweet texts was critical to reduce noise and standardize the input for sentiment analysis. A custom preprocessing pipeline was implemented using NLTK's WordNetLemmatizer and stopword corpus. The pipeline executed the following steps sequentially: removal of URLs using regular expressions (re.sub(r'http\S+|www\S+|https\S+', '', text)), stripping of user mentions (@\w+), retention of hashtag keywords after removing the '#' symbol, and elimination of punctuation characters using Python's str.translate method with string.punctuation. All text was converted to lowercase, numeric characters were removed (re.sub(r'\d+', '', text)), and tokenized via simple whitespace splitting. Tokens shorter than two characters and those appearing in an extended stopword set—including domain-specific words such as "afghan," "women," "taliban," and "amp"—were discarded. Lemmatization reduced tokens to their canonical form, improving term matching across tweets. The cleaned tokens were rejoined into processed sentences, which were stored in a new column cleaned_content. Tweets that resulted in empty cleaned text after preprocessing were removed to maintain data integrity. To visually summarize frequent terms post-cleaning, a word cloud was generated with a maximum of 200 words, excluding common collocations to better highlight distinct thematic vocabulary.

Sentiment Annotation and Feature Extraction

Sentiment labels were assigned automatically using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool from NLTK, which is optimized for social media text. The compound sentiment score thresholding was set as follows: scores ≥ 0.05 indicated positive sentiment, ≤ -0.05 indicated negative sentiment, and values between these thresholds were labeled neutral. This generated three sentiment classes used as target labels for supervised learning. Text features were extracted using the TF-IDF (Term Frequency–Inverse Document Frequency) vectorizer from scikit-learn, configured with a maximum vocabulary size of 5,000 features, and an n-gram range of 1 to 2 (unigrams and bigrams). Terms appearing in fewer than two documents (min_df=2) were excluded to reduce noise and dimensionality. The vectorizer transformed the cleaned text into a sparse matrix representation suitable for model training.

Model Training and Evaluation

The dataset was split into training and testing subsets with an 80:20 ratio using train_test_split, applying stratification based on sentiment labels to preserve class distribution. In cases where any sentiment class had fewer than two samples, stratification was disabled to avoid splitting errors. A Multinomial Naive Bayes classifier was selected due to its suitability for discrete feature spaces such as TF-IDF and its proven effectiveness in text classification. The model was trained on the training set with default hyperparameters (e.g., smoothing parameter alpha=1.0). After training, the model's performance was evaluated on the test set using multiple metrics: accuracy, precision, recall, and F1-score, all computed via scikit-learn’s classification_report. A confusion matrix was generated to visualize the true versus predicted sentiment classes, plotted using seaborn’s heatmap for clear interpretability. All evaluation artifacts—including model checkpoints, vectorizer objects, and evaluation results—were saved for reproducibility.

Feature Importance and Temporal Sentiment Analysis

To interpret the linguistic drivers behind the model’s sentiment classifications, the top 20 features contributing most strongly to each sentiment class were identified by sorting the log probabilities (feature_log_prob_) learned by the Naive Bayes model. These features were visualized as horizontal bar charts to demonstrate which terms carried the highest weight for positive, negative, and neutral sentiment predictions. Additionally, sentiment trends over time were analyzed by grouping the labeled tweets by month and calculating the monthly distribution of each sentiment category. Line plots depicted the proportion of positive, negative, and neutral tweets per month, revealing temporal shifts in public opinion. These visualizations provided valuable insights into how global sentiment regarding Afghan women's rights evolved in response to real-world events during the studied period.

Result and Discussion

Data Overview and Exploratory Data Analysis

The initial dataset comprised 5,000 tweets, each with 20 columns representing various metadata and content attributes. After inspection, key findings included the presence of some missing values, notably 100% missing in the retweetedTweet column and significant gaps in coordinates (98.5%) and quotedTweet (86.16%). Despite these gaps, the dataset was largely complete for critical fields like content, date, and engagement metrics. Filtering for English-language tweets reduced the dataset slightly to 4,862 entries, focusing the analysis on tweets accessible by the English-based tools employed. The dataset's temporal coverage spanned the intended period from April 2023 to January 2024.

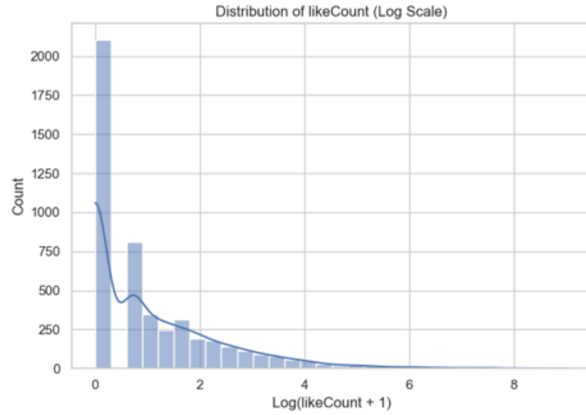

Figure 1 Distribution of Like Count

Figure 1 shows the distribution of likes received by the English tweets, plotted on a logarithmic scale (Log(Likes + 1)) due to the skewed nature of engagement data (many tweets get few likes, a few get many). The x-axis is the log-transformed like count, and the y-axis is the frequency of tweets. The distribution appears right-skewed even on the log scale, indicating most tweets receive a low number of likes.

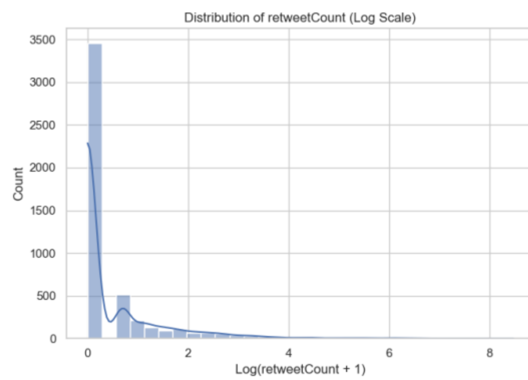

Figure 2 Distribution of Retweet Count

Figure 2 displays the distribution of retweets on a logarithmic scale (Log(Retweets + 1)). The shape is also right-skewed, showing that a vast majority of tweets have very few or no retweets, while a small number of tweets achieve higher retweet counts.

Text Preprocessing and Sentiment Distribution

After cleaning and preprocessing the tweet texts—removing URLs, mentions, stopwords, punctuation, and performing lemmatization—the final corpus consisted of 4,845 tweets. Only 17 tweets were discarded during preprocessing due to resulting empty content, indicating effective text retention. The generated word cloud visually emphasized common thematic terms, confirming the focus on Afghan women and related human rights issues. Sentiment labeling using VADER (figure 3) showed that negative sentiment was the most prevalent, accounting for 2,297 tweets (47.4%), followed by positive sentiment at 1,857 tweets (38.3%) and neutral sentiment at 691 tweets (14.3%). This distribution highlights the global concern and emotional response to the plight of Afghan women reflected in social media discourse.

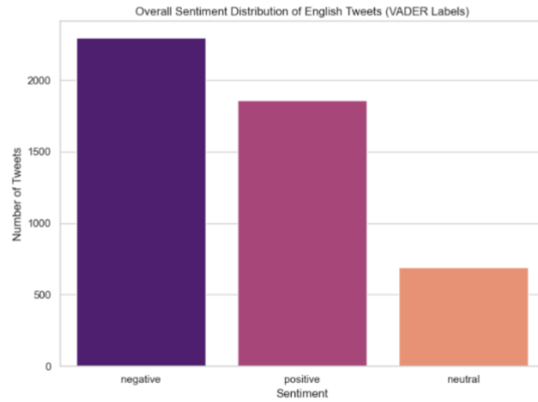

Figure 3 **Sentiment Distribution (VADER Labels)**

Model Training and Performance Evaluation

Using the cleaned and labeled data, a TF-IDF vectorizer transformed the text into a sparse feature matrix of shape (4,845, 5,000), representing the frequency and importance of terms across tweets. The dataset was split into training (3,876 tweets) and testing (969 tweets) subsets using stratified sampling to maintain sentiment class proportions. A Multinomial Naive Bayes classifier was trained with default hyperparameters. On the test set, the model achieved an overall accuracy of approximately 67.5%, demonstrating moderate effectiveness in predicting sentiment based on textual features. Precision and recall metrics varied by class: negative tweets achieved 0.66 precision and 0.87 recall, indicating the model was better at identifying true negatives but with some false positives. Positive tweets had balanced precision and recall of around 0.69 and 0.64 respectively. Neutral tweets were challenging, with high precision (0.94) but very low recall (0.12), suggesting the model rarely missed neutral tweets it predicted but failed to identify most neutral tweets present.

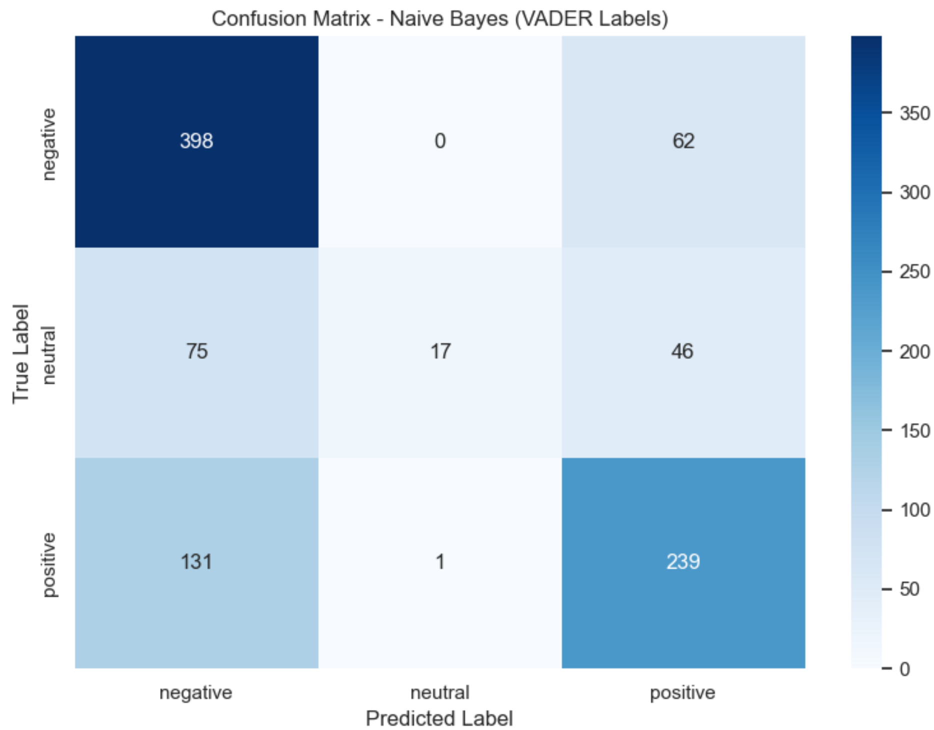

Figure 4 Confusion Matrix

The confusion matrix (figure 4) visualized these results, showing a strong diagonal for negative and positive classes but more misclassifications involving the neutral class. The weighted F1-score was 0.64, reflecting the overall balance of precision and recall. These metrics confirm that while the model can generally distinguish sentiment polarity, neutral sentiment remains difficult to capture accurately, possibly due to its subtler linguistic cues.

Feature Importance Analysis

Analysis of feature importance revealed the top words influencing sentiment classification for each class. For the negative class, terms strongly associated with adverse sentiment included words like “ban,” “kill,” and “force,” reflecting concerns over restrictions and violence. Positive sentiment was linked to words such as “support,” “hope,” and “freedom,” indicating expressions of solidarity and optimism. Neutral features generally contained more generic or factual terms. Visualizations of these feature importance rankings were saved as bar charts, providing interpretable insights into the model’s decision-making process and highlighting the vocabulary most indicative of public sentiment on Afghan women’s rights.

Sentiment Trends Over Time

Monthly sentiment trends illustrated dynamic changes in public discourse. The proportion of negative sentiment dominated throughout the period but showed fluctuations with occasional rises in positive and neutral tweets. These trends corresponded to specific events and announcements related to Afghan women’s rights, indicating that real-world developments influenced online emotional responses. The visualization of sentiment proportions over time allowed for the identification of key moments of heightened engagement and shifts in public opinion. Such temporal analysis underscores the value of social media as a real-time barometer of societal attitudes on critical human rights issues.

Discussion

The dominance of negative sentiment in the tweets analyzed reflects a profound and widespread international concern regarding the rights and well-being of Afghan women. This pervasive negativity suggests that many individuals around the world perceive the situation as dire, marked by restrictions, violence, and significant violations of basic human rights. Such a strong emotional response signals that the issue resonates deeply across global digital communities, highlighting the urgency and gravity of the ongoing challenges faced by Afghan women.

These findings highlight the critical role social media platforms, especially Twitter, play in bringing human rights issues to the forefront of public consciousness. Through rapid sharing and amplification, social media enables activists, journalists, and ordinary users to raise awareness about injustices that might otherwise remain underreported. The volume and tone of the discourse reveal how digital platforms function as powerful tools for advocacy, capable of mobilizing international support and potentially influencing policy decisions and humanitarian actions.

When compared to sentiment analyses conducted on broader human rights topics, the discourse around Afghan women's rights appears significantly more polarized. Whereas other human rights discussions may show more balanced or neutral sentiment distributions, the strong tilt toward negative and, to a lesser extent, positive sentiment in this study points to a global call for urgent action. This polarization reflects the heightened emotional stakes and the intense spotlight on Afghanistan's political and social upheaval, distinguishing it from other issues in the digital human rights landscape.

However, the study is not without limitations. The exclusive reliance on Twitter data constrains the generalizability of the results, as different social media platforms attract diverse user demographics and engagement patterns. For example, platforms like Facebook, Instagram, or regional networks may host conversations with different characteristics, possibly altering the sentiment landscape. Therefore, the findings represent only a partial view of the broader digital discourse.

Another limitation stems from the focus on English-language tweets. While this choice aligns with the analytical tools used, it excludes a potentially substantial body of content in languages such as Pashto, Dari, Arabic, and others spoken by key regional stakeholders. This linguistic limitation may lead to the underrepresentation of perspectives from within Afghanistan and neighboring countries, which could provide valuable cultural and contextual insights.

Despite these limitations, the study offers significant insights into the nature of international digital discourse on Afghan women's rights. It emphasizes the value of social media sentiment analysis as a lens through which global public opinion can be understood and monitored. Future research expanding across multiple platforms and languages could build upon these findings to develop a more comprehensive understanding of the digital activism landscape surrounding human rights in conflict zones.

Conclusion

The sentiment analysis conducted on tweets concerning Afghan women's rights highlights a significant and widespread global concern. The predominance of negative sentiment reflects international frustration and alarm over the ongoing violations and restrictions faced by women in Afghanistan. These findings underscore the emotional weight and urgency surrounding this human rights issue as expressed through social media discourse.

This study contributes to the growing body of research that demonstrates the power of social media platforms in both mirroring and shaping public opinion on critical human rights matters, especially within politically unstable and conflict-affected regions. By focusing on Twitter data, the research reveals how digital conversations can provide timely insights into global perceptions, advocacy efforts, and the mobilization of support for vulnerable populations.

Looking ahead, future research can build on this foundation by broadening the scope to include multiple social media platforms such as Facebook and Instagram, which may capture a more diverse and comprehensive range of voices. Additionally, incorporating tweets and posts in various languages would enrich the understanding of regional and cultural perspectives. Further investigations could also examine how public sentiment evolves in response to international interventions, policy changes, or major events related to Afghan women's rights, offering valuable guidance for activists and policymakers.

Declarations

Author Contributions

Conceptualization: T.S.; Methodology: L.Y.; Software: L.Y.; Validation: L.Y.; Formal Analysis: T.S.; Investigation: T.S.; Resources: L.Y.; Data Curation: T.S.; Writing Original Draft Preparation: L.Y.; Writing Review and Editing: L.Y.; Visualization: T.S.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Keith, “The Impact of Social Media on Political Activism,” *Ijhss*, vol. 1, no. 1, pp. 16–29, 2023, doi: 10.47941/ijhss.1429.

[2] Y. AM *et al.*, “High Social Media Use Among Adolescents Associated With Increased High-Risk Behaviors and Poor Academic Outcomes,” *J. Fam. Med.*, no. 6, 2022, doi: 10.26420/jfammed.2022.1310.

[3] K. E. Riehm, J. Thrul, J. L. Barrington-Trimis, A. Kelleghan, R. Mojtabai, and A. M. Leventhal, “Prospective Association of Digital Media Use With Alcohol Use Initiation and Progression Among Adolescents,” *Alcohol. Clin. Exp. Res.*, vol. 45, no. 4, pp. 877–885, 2021, doi: 10.1111/acer.14578.

[4] B. C. McHugh, P. Wiśniewski, M. B. Rosson, and J. M. Carroll, “When Social Media Traumatizes Teens,” *Internet Res.*, vol. 28, no. 5, pp. 1169–1188, 2018, doi: 10.1108/intr-02-2017-0077.

[5] N. A. Santoso, . Alwiyah, and E. A. Nabila, “Social Media Factors and Teen Gadget Addiction Factors in Indonesia,” *Adi J. Recent Innov. Ajri*, vol. 3, no. 1, pp. 67–77, 2021, doi: 10.34306/ajri.v3i1.289.

[6] D. Taibi, “Social Media Literacy to Support a Conscious Use of Social Media in Adolescents and Improve Their Psychological Well-Being: A Pilot Study,” *Sustainability*, vol. 15, no. 17, p. 12726, 2023, doi: 10.3390/su151712726.

[7] E. B. Tuğtekin and M. Koç, “Understanding the Relationship Between New Media Literacy, Communication Skills, and Democratic Tendency: Model Development and Testing,” *New Media Soc.*, vol. 22, no. 10, pp. 1922–1941, 2019, doi: 10.1177/1461444819887705.

[8] N. Suzor, M. Dragiewicz, B. Harris, R. Gillett, J. Burgess, and T. V. Geelen, “Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online,” *Policy Internet*, vol. 11, no. 1, pp. 84–103, 2018, doi: 10.1002/poi3.185.

[9] H. Al-Dmour, R. Masa’deh, A. Salman, M. Abuhashesh, and R. Al-Dmour, “Influence of Social Media Platforms on Public Health Protection Against the COVID-19 Pandemic via the Mediating Effects of Public Health Awareness and Behavioral Changes: Integrated Model,” *J. Med. Internet Res.*, vol. 22, no. 8, p. e19996, 2020, doi: 10.2196/19996.

[10] K. M. Ashifa, “Human Right Awareness and Advocacy Role of Youth in Kerala: An Empirical Analysis,” *Rupkatha J. Interdiscip. Stud. Humanit.*, vol. 12, no. 1, 2020, doi: 10.21659/rupkatha.v12n1.27.

[11] A. Jungherr, H. Schoen, O. Posegga, and P. Jürgens, “Digital Trace Data in the Study of Public Opinion,” *Soc. Sci. Comput. Rev.*, vol. 35, no. 3, pp. 336–356, 2016, doi: 10.1177/0894439316631043.

[12] J. Mellon and C. Prosser, “Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users,” *Res. Polit.*, vol. 4, no. 3, 2017, doi: 10.1177/2053168017720008.

[13] Y.-S. Hwang, “Does Opinion Leadership Increase the Followers on Twitter,” *Int. J. Soc. Sci. Humanity*, vol. 5, no. 3, pp. 258–264, 2015, doi: 10.7763/ijssh.2015.v5.464.

[14] A. Elyashar, I. Plochotnikov, I.-C. Cohen, R. Puzis, and O. Cohen, “The State of Mind of Health Care Professionals in Light of the COVID-19 Pandemic: Text Analysis Study of Twitter Discourses,” *J. Med. Internet Res.*, vol. 23, no. 10, p. e30217, 2021, doi: 10.2196/30217.

[15] S. Boon-itt and Y. Skunkan, “Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study,” *Jmir Public Health Surveill.*, vol. 6, no. 4, p. e21978, 2020, doi: 10.2196/21978.

[16] Z. Reyhana, K. Fithriasari, M. Atok, and N. Iriawan, “Linking Twitter Sentiment Knowledge With Infrastructure Development,” *Matematika*, pp. 91–102, 2018, doi: 10.11113/matematika.v34.n3.1142.

[17] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, “The State-of-the-Art in Twitter Sentiment Analysis,” *Acm Trans. Manag. Inf. Syst.*, vol. 9, no. 2, pp. 1–29, 2018,

doi: 10.1145/3185045.

[18] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment in Twitter Events,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, no. 2, pp. 406–418, 2010, doi: 10.1002/asi.21462.

[19] D. E. Cahyani and I. Patasik, “Performance Comparison of TF-IDF and Word2Vec Models for Emotion Text Classification,” *Bull. Electr. Eng. Inform.*, vol. 10, no. 5, pp. 2780–2788, 2021, doi: 10.11591/eei.v10i5.3157.

[20] B. Pahwa, S. Taruna, and N. Kasliwal, “Sentiment Analysis- Strategy for Text Pre-Processing,” *Int. J. Comput. Appl.*, vol. 180, no. 34, pp. 15–18, 2018, doi: 10.5120/ijca2018916865.

[21] A. E. S. Saputro, K. A. Notodiputro, and A. Indahwati, “Study of Sentiment of Governor’s Election Opinion in 2018,” *Int. J. Sci. Res. Sci. Eng. Technol.*, pp. 231–238, 2018, doi: 10.32628/ijsrset21841124.

[22] A. Mustolih, “Sentiment Analysis Motorku X Using Applications Naive Bayes Classifier Method,” *Indones. J. Artif. Intell. Data Min.*, vol. 6, no. 2, p. 231, 2023, doi: 10.24014/ijaidm.v6i2.24864.

[23] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, “Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 1, p. 57, 2018, doi: 10.20473/jisebi.4.1.57-64.

[24] N. Rezki, S. A. Thamrin, and S. Siswanto, “Sentiment Analysis of Merdeka Belajar Kampus Merdeka Policy Using Support Vector Machine With Word2vec,” *Barekeng J. Ilmu Mat. Dan Terap.*, vol. 17, no. 1, pp. 0481–0486, 2023, doi: 10.30598/barekengvol17iss1pp0481-0486.

[25] A. Chakravarthy, “Hybrid Architecture for Sentiment Analysis Using Deep Learning,” *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 1, pp. 735–738, 2018, doi: 10.26483/ijarcs.v9i1.5388.

[26] N. F. Tan, “Group-Based Protection of Afghan Women and Girls Under the 1951 Refugee Convention,” *Int. Comp. Law Q.*, vol. 72, no. 3, pp. 793–817, 2023, doi: 10.1017/s0020589323000283.

[27] B. Lalyzada, “Role of Media and Social Networks in Promoting the Position of Afghan Women’,” *Sprin J. Arts Humanit. Soc. Sci.*, vol. 3, no. 1, pp. 48–56, 2024, doi: 10.55559/sjahss.v3i1.227.

[28] B. G. Khalili, “Assessing the Impact of Social Media on Youth’s Entrepreneurship Development in Afghanistan,” *Ijist*, vol. 2, no. 5, pp. 396–411, 2024, doi: 10.59890/ijist.v2i5.1754.

[29] K. Becvar, “The ‘First Daughter’ Effect: Human Rights Advocacy and Attitudes Toward Gender Equality in Taliban-Controlled Afghanistan,” *Plos One*, vol. 19, no. 7, p. e0298812, 2024, doi: 10.1371/journal.pone.0298812.

[30] T. T. Nguyen *et al.*, “Pride, Love, and Twitter Rants: Combining Machine Learning and Qualitative Techniques to Understand What Our Tweets Reveal About Race in the US,” *Int. J. Environ. Res. Public. Health*, vol. 16, no. 10, p. 1766, 2019, doi: 10.3390/ijerph16101766.

[31] L. L. Manganaro and N. O. Alozie, “Gender Role Attitudes: Who Supports Expanded Rights for Women in Afghanistan?,” *Sex Roles*, vol. 64, no. 7–8, pp. 516–529, 2011, doi: 10.1007/s11199-011-9931-6.

[32] R. Afrouz, B. R. Crisp, and A. Taket, “Afghan Women Perceptions of Gender Roles, Possibilities and Barriers to Change After Settlement in Australia: A Qualitative Study,” *Qual. Soc. Work*, vol. 22, no. 3, pp. 569–586, 2022, doi: 10.1177/14733250221076730.

[33] J. Jagarapu, “Twitter Discussions on Breastfeeding During the COVID-19 Pandemic,” *Int. Breastfeed. J.*, vol. 18, no. 1, 2023, doi: 10.1186/s13006-023-00593-x.