Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 2, May 2026, pp. 584~595
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i2.pp584-595

A multi-class classification approach for feminist sentiment analysis in Bangla social media using TF-IDF and ensemble learning

Zaid Bin Sajid, Md. Mijanur Rahman, Md. Sumon Hosen, Sarara Jaman Riya, Yeamin Akon, S. M. Fahad Bin Jim, Ornab Biswass
Department of Computer Science and Engineering, School of Engineering, Southeast University, Dhaka, Bangladesh

Article Info

Article history:

Received Dec 15, 2025
Revised Mar 24, 2026
Accepted May 26, 2026

Keywords:

Abusive language detection
Bangla NLP
Bangla sentiment
Feminism and hate speech
Machine learning

ABSTRACT

Social media has emerged as an important part of societal discourse on feminism and gender equality, especially in Bangladesh. Nevertheless, any feminist debate on social media in Bengali polarizes reactions, highlighting the need for automated sentiment analysis. This paper introduces one of the earliest multi-class feminist sentiment classification schemes of the Bengali social media with a manually annotated dataset of 6,830 comments categorized as positive, neutral, or negative. The framework uses term frequency-inverse document frequency (TF-IDF) based n-gram feature representations utilizing traditional machine learning algorithms, with a majority voting ensemble to determine optimal robust models. The data was divided into 80% and 20% for training and testing, respectively. Models were evaluated on the basis of accuracy, precision, recall, and macro-F1 to correct on imbalance of classes. Multinomial naive bayes (MNB) has the best accuracy of 84.74% and macro-F1 of 84.66, which is 4-7 times higher than other models. The ensemble method improved feature strength. Such results indicate that lightweight machine learning models based on TF-IDF features and ensemble models can be useful to detect feminist sentiment in Bangla social media and serve as a guideline in the field of domain-specific sentiment analysis in low-resource languages and help monitor online feminist discourse.

Corresponding Author:

Md. Mijanur Rahman
Department of Computer Science and Engineering, School of Engineering, Southeast University
251/A Tejgaon I/A, Dhaka 1208, Bangladesh
Email: mijanur.rahman@seu.edu.bd

1. INTRODUCTION

The recent spread of social sites has offered new opportunities to people to voice their complaints publicly and engage in societal debate on different topics. Social media is seizing the social conversation, and Bangladesh, Dhaka in particular, is remarkably high in its usage [1], [2]. Technological advances, on the other hand, have been used to spread false information, encourage hate speech, and attack people based on their ethnicity, religion, gender, or other distinctions [3]. Bias argumentation against certain groups is also rife [4]. Detection of sentiment focused on feminism is an area that has received more and more attention in natural language processing, especially hate speech. Past studies have utilized many different methodologies, such as rule-based systems, automated learning algorithms, support vector machines (SVM), K-nearest neighbors (KNN), and, more recently, deep learning models [5], [6]. Modifications of the detection of misogynistic text in the Bengali language using a transformer-based model have also been reported in

associated literature, although in the context of YouTube commentaries as the new landscape of misogyny detection within social media networks [7]. Furthermore, transformer-based models like BanglaBERT have been employed to boost multiclass hate speech detection in the Bengali social media [8].

More recent work on Bangla multimodal and transformer-based learning has also demonstrated that deep learning architectures can be helpful to complete multiclass classification tasks in crisis and social media spaces [9]. A general overview of hate speech detection in the Bengali language has also made known the gains and the challenges faced in this arena as a subset of the general picture of abusive language detection [10].

Whereas studies of abusive language detection may be found in English, languages like Bangla lack a resource shortage [11]-[13]. The complexity of the linguistic and cultural context, sarcasm, and colloquialism in social media makes the Bengali language a complex problem that necessitates a tailored approach to accurate sentiment analysis and abuse detection [9], [14]. The use of Bangla is widespread, and approximately 290 million people in the world identify with it as their original language [15]. Feminist voices in Bangladesh are frequently met with hate and abuse on the internet, which forms a big hindrance to their participation and action. Recent studies have discussed the analysis of abusive remarks against feminism using machine-learning techniques [16], [17]. Regardless of these advances, the literature on the subject is overwhelmed by general or small-scale datasets and does not extend into further areas of feminist discourse because of the application of multi-class abusive language detection in Bengali. A number of papers have examined emotional scoring and hate speech recognition in Bangla [18]. They often display ignorance of socio-cultural context, especially in feminist discourse. Although there is an increasing amount of literature on Bangla hate speech detection, the available research mostly resorts to general-purpose or binary classification datasets and fails to explicitly contrast feminist-oriented discourse with multi-class abusive language detection, which creates a considerable gap in the research domain with specific and context-dependent modeling. This paper seeks to fill this gap by coming up with a more practical approach to sentiment analysis in Bangla feminist literature using machine learning.

Although the previous data article settled on dataset building, annotation, and ethical aspects mostly [19], the current study builds on the previous one by implementing feature engineering and machine learning methods to identify abusive language in the Bangla feminist discourse. One of the aims is to create and train a machine learning model that can categorize user comments with accuracy in sentiment, thus helping the identification of positive, negative, and neutral language in texts on feminism in Bengali. Specifically, the study makes several significant contributions: Gathering a large amount of data on the topic and facilitating the analysis of public opinions on feminism, we present a manually annotated feminist-oriented Bengali dataset. The paper will be structured in the following way: section 1 will contain the introduction and a comprehensive literature review of the existing literature. Section 2 presents the methodology of this research, which consists of data collection, annotation, preprocessing, feature engineering, model training, and evaluation. Section 3 includes the research results and gives an account of the model's performance. The discussion is in section 4, and for the possible future work, we recap our work conclusion in section 5.

Offensive language has also been studied by researchers in recent years through the policies of machine learning (ML) and deep learning (DL), including the Bengali language [20], [21]. To illustrate, Levchenko and Dilai *et al.* [22] conducted a study of Ukrainian feminist tweets with the help of the SentiStrength algorithm to assess the sentiment. However, the bulk of earlier studies has been defined by small datasets and poor performances, rendering generalization of such studies hard. Scarborough [23] applied qualitative and naive Bayes sentiment analysis algorithms to gauge the opinion of the people on gender in 9 regions in the USA and obtained an accuracy of 71%. He used a large pool of tweets, but it was not representative of the gender attitudes of nonwhite groups and the less-educated individuals/individuals with limited formal education. Albadi *et al.* [24] have created a publicly available Arabic dataset that is designed to identify religious hate speech and has offered a variety of classification models. Their accuracy was 79%, but it was a small dataset that the authors provided: 1000 tweets per six religions.

Moreover, Herwanto *et al.* [20] trained a hate speech classification model based on the fastText algorithm and word representation in the form of a continuous bag-of-words (CBOW) model. This study will help to guide Indonesia in detecting hate speech on social media early enough to avert misunderstanding and criminal offenses. The model performs well under binary classification, though not significantly better than the previous studies. The accuracy is 65.71%. The size of the dataset is 700 comments, which is rather restricted.

A 1141-data-points model proposed by Hassan *et al.* [25] was used to take the sentiment, either positive or negative, based on the chat. The dataset was used with SVM, multinomial naive bayes (MNB), KNN, logistic regression (LR), decision tree (DT), and random forest (RF). The best accuracy was 86 percent by SVM. Emon *et al.* [2] identified various categories of abusive language online using deep learning algorithms, which are most effective with 82.20% accuracy in recurrent neural network (RNN). The Bangla language has also been implemented with new stemming rules, which have resulted in improved performance

of the algorithm. They applied Linear SVM, LR, MNB, RF, artificial neural network (ANN), RNN, and long short-term memory (LSTM) to the 4700 dataset. Tusher *et al.* [26] applied ML techniques to study the addiction to internet gaming to over 401 data points. To train the suggested algorithm and detect three groups of words in the Bangla language in positive, negative, and neutral data, researchers use six machine learning algorithms: DT, RF, MNB, extreme gradient boosting (XGB), SVM, and KNN. The maximum accuracy was 73.27 percent by MNB, though the authors utilized certain particular data sets. Hasan *et al.* [11] examined various publicly available datasets with sentiment labels and prepared classifiers based on conventional and deep learning methods. They employed the SVM, RF, convolutional neural network (CNN), FastText, and transformer-based models, of which they achieved the greatest accuracy of 61.2%. However, this was work conducted on the low-resource languages, the existing literature on Bangla, and the annotated datasets.

Tuhin *et al.* [27] employed a topical approach and naive Bayes classification algorithm to six different categories of emotion categorization, which are happy, sad, tender, enthusiastic, angry, and afraid, in the context of sentiment and emotion analysis of Bangla textual data; the study resulted in 90%. Existing sources on the topic of Bangla studies are largely focused on general hate speech or binary classification, and the idea of feminist-based datasets and the multi-class abusive language detector are under-researched. Nevertheless, the author might have adopted a more powerful algorithm to ensure excellent performance.

2. METHOD

The methodology encompasses data collection, annotation, dataset preprocessing, relevant feature extraction, model development, and performance evaluation. The proposed approach starts with data collection, preprocessing, and analyzing data to uncover patterns of abuse in Bangla-language comments. The complexities of natural language processing (NLP) in Bangla, a language with its unique linguistic structure and nuances. This study carefully considers the specific challenges posed by the language. The methodology is organized into key phrases, each contributing to the study's overall goal. For the model evaluation, we have applied an ensemble technique. Since ensemble ML is thought to be the most advanced answer for many machine learning issues, data science professionals in several industries employ this powerful machine learning technique [28]. Figure 1 visually represents the workflow, providing a clear overview of the steps involved.

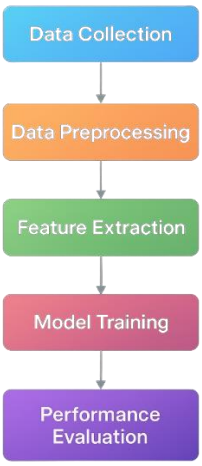


Figure 1. Diagram of our proposed methodology

2.1. Data Collection

Data collection serves as the initial and fundamental step of this study. This study has not used a publicly available dataset; rather, all data have been collected manually, where relevant comments are gathered from social platforms (e.g., Facebook, YouTube, and Instagram). Comments related to feminist topics using keywords and hashtags such as "নারীবাদ" (feminism), "নারীবাদী" (feminist), and "নারীর অধিকার" (women's rights) were used to filter the data. In our experiment, we gathered 6,830 datasets manually, as mentioned in Table 1. Both Bangla and English comments are included; afterwards, English comments are translated into Bangla using Google Translate to maintain linguistic consistency. The dataset was collected from Bangla social media platforms to capture real-world linguistic variability, including informal usage, slang, and noisy user-generated content.

Table 1. Count of comments

Label	Number of comments
Positive	2,346
Negative	2,294
Neutral	2,190

2.1.1 Sentiment annotation

Sentiment annotation involves labeling the collected comments with sentiment categories such as positive, neutral, and negative. This phase may involve both manual and automatic annotation methods. However, this study involves manual annotation, and we have labeled these data. Data annotation is a challenging technique due to sarcasm and colloquial language. Each comment was painstakingly annotated by native Bangla speakers to make sure it was relevant and to spot any offensive language. This manual validation process ensures a high-quality dataset suitable for machine learning and natural language processing tasks, such as sentiment analysis, abusive language analysis, and the study of online gender-based violence in the Bangla language context. A representative example from the dataset is presented in Figure 2, Figure 3, and Figure 4 represent the distribution of our dataset.

	Comment	Label
0	সহবাস কইরা দিলা খালাম্মারে	Negative
1	খুব ভাল লাগল আপু আপনার কথা গুলা শুনে	Positive
2	সাহস থাকলে একদিন মেকআপ ছাড়া সামনে আসো তো?	Negative
3	ভাই মালটাকে নাইট ক্লাবে নিয়ে যান একটু😡	Negative
4	আল্লাহ তা,আলা আমাদের যে হুকুম করেছেন সেইভাবেই ...	Neutral
5	খুব সুন্দর 👍	Positive
6	তার লিখিত বই গুলো তোমার মা মেয়ে ও বোন কে পড়তে দিও	Neutral
7	ওহ। আরেকজন খদ্দের পাওয়া গেল।	Negative
8	মাশা-আল্লাহ, লেখাটা পড়ে ভালো লাগলো।	Positive
9	দেখবে কেউ না কেউ ওদের পক্ষে দাঁড়িয়ে যাবে!আমি...	Neutral

Figure 2. Manual dataset annotation

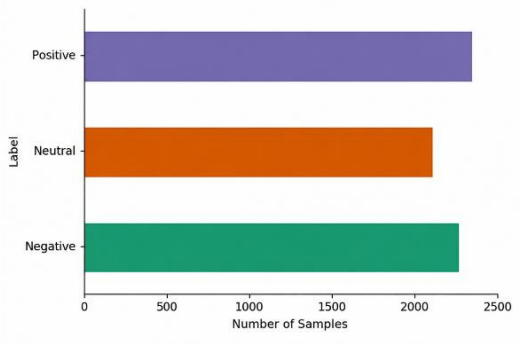


Figure 3. Dataset distribution

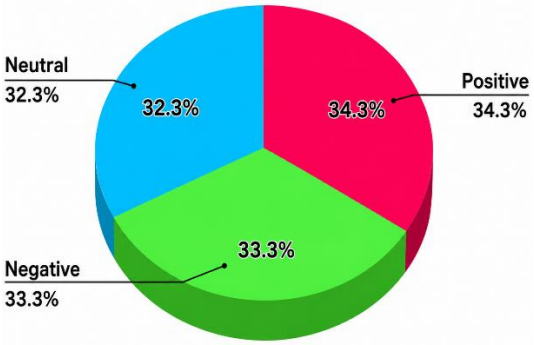


Figure 4. Percentage distribution

2.2. Data preprocessing

Preprocessing techniques were used to reduce the noise that is frequently present in Bangla social media text, such as emojis, short word usage, and special characters.

2.2.1. Dataset cleaning

Dataset cleaning is an essential preprocessing step that directly impacts the model's performance. User-generated content, especially in Bangla, often contains various forms of noise and irregularities, irrelevant information, and inconsistencies. User comments often contain special characters like #, $, @, %, punctuation marks, emoji, unnecessary spaces, and numeric digits that do not contribute to understanding the sentiment or abusive content. These elements are removed to simplify the text and focus on meaningful content. The following Figure 5 represents the original and cleaned text.

Original	Cleaned
ভাই আপনিতো অমানুষ,,,এখনও মানুষ হননাই তাইনা,,,তাহলে অমানুষকে কি বুঝবো,,,,,কথাটা কষ্ট লাগলো বুঝি,,,কারন আপনি নিজেই আপনার বায়োতে লৈইখা রাখছেন, যেয়ে দেখেন মিয়া,,,,বলদের মতো কথাবার্তা, এমনে লেখতেন,,খারাপ আমি ভালো হওয়ার চেষ্টা করতাছি,,,,,😭😭😭	ভাই আপনিতো অমানুষ এখনও মানুষ হননাই তাইনা তাহলে অমানুষকে কি বুঝবো কথাটা কষ্ট লাগলো বুঝি কারন আপনি নিজেই আপনার বায়োতে লৈইখা রাখছেন যেয়ে দেখেন মিয়া বলদের মতো কথাবার্তা এমনে লেখতেন খারাপ আমি ভালো হওয়ার চেষ্টা করতাছি
100% ঠিক। একজন নারী হয়ে সত্য কথা বলার জন্য আপনাকে অসংখ্য ধন্যবাদ।	ঠিক একজন নারী হয়ে সত্য কথা বলার জন্য আপনাকে অসংখ্য ধন্যবাদ
এই তোকে কি জিজ্ঞেস করেছি আর তুই কি উত্তর দিয়েছিস? জন্মের দোষ টোষ আছে নাকি তোর? নাকি পরিবার থেকে কিভাবে মানুষের সাথে আলাপচারিতা করতে তার শিক্ষা দেয়নি?	এই তোকে কি জিজ্ঞেস করেছি আর তুই কি উত্তর দিয়েছিস জন্মের দোষ টোষ আছে নাকি তোর নাকি পরিবার থেকে কিভাবে মানুষের সাথে আলাপচারিতা করতে তার শিক্ষা দেয়নি
মন্তব্যটি অনেক কিছু ভাবতে সাহায্য করেছে।	মন্তব্যটি অনেক কিছু ভাবতে সাহায্য করেছে
সালা কিছু না জেনে ধন্যবাদ চোদাও কেন???	সালা কিছু না জেনে ধন্যবাদ চোদাও কেন

Figure 5. Examples of original and cleaned text after preprocessing

2.2.2. Remove low length comments

Short comments that are too brief, such as single-word or one-character entries (e.g., “হ্যাঁ”, “না”, “ওকে”), often provide little meaningful information. These low-length comments can introduce noise or be ambiguous, so we removed comments below a defined threshold. In our dataset, we identified 18 comments with a length of 1 character that were deemed too short to be helpful for the task at hand. After removing these low-length comments, we are left with 6,812 comments for further processing.

2.2.3. Data visualization

We utilize word cloud as a data visualization technique to gain a deeper understanding of the cleaned dataset. A word cloud visually illustrates the words that appear most frequently in the dataset, where the dimension of each word indicates how frequently it occurs. This helps identify common themes and patterns in the comments, offering insight into which words are most prominent across positive, neutral, and negative comments that have been depicted in Figure 6.

Figure 6. Most frequently appearing words in the cleaned dataset

2.2.4. Encoding labels and dataset splitting

After text cleaning, labels (e.g., positive, negative, neutral) are converted into numeric values, a necessary step for training machine learning models. For instance, positive is encoded as 2, negative as 0, and neutral as 1. This encoding enhances the model’s understanding of the relationships between the text and its sentiment in a structured manner, enabling effective classification. Subsequently, the refined dataset is divided into training and testing sets, where 80% is utilized for training and 20% is kept for testing purposes. For our dataset of 6,812 (after cleaning), the split results were 5,449 comments for training and 1,363 comments for testing. This balanced division ensures the model has sufficient data to learn from while rigorously evaluating diverse examples during testing that shows in Figure 7.

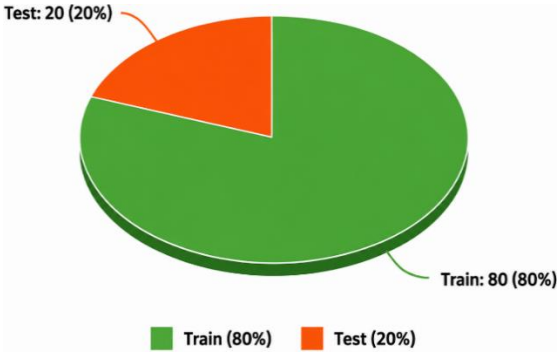


Figure 7. Dataset information

2.3. Feature extraction

For effective representation of textual data, we employed term frequency-inverse document frequency (TF-IDF) weighting together with Unigram, Bigram, and Trigram approaches. These established approaches in NLP help extract features by quantifying the significance of terms within individual documents in relation to their overall presence in the corpus. The feature sizes, as shown in Table 2, were determined after applying TF-IDF to the preprocessed dataset. Each document in the corpus was represented as a sparse vector corresponding to the respective n-gram feature. These feature sizes represent the dimensionality of the input to our machine-learning models during training.

Table 2. N-gram feature size

N-gram type	Feature size
Unigram	12,592
Bigram	73,387
Trigram	150,334

2.4. Model training

After the transformation of a textual data into feature vectors, several machine learning models were trained. Various standard algorithms such as MNB, LR, RF, KNN, DT, LSVM, and KSVM were implemented in the study. All models were trained on TF-IDF feature vectors created on unigram, bigram and trigram representations.

- MNB: MNB is a standard probabilistic learning method used when dealing with text classification. The features (typically the word frequencies) are assumed to be conditionally independent of the class and of a multinomial form [29].
- LR: there is the use of logistic (sigmoid) in the linearization phase to scale outputs to probabilities, which is called LR to model the probability of membership in a class. In text categorization, it is effective when the data is linearly separable [30].
- RF: the RF ensemble learning method is applied to first create a set of DTs and the output of these DTs is then joined together to create a reliable classification. The method works well in increasing generalization and minimizing overfitting particularly in noisy datasets. It is effective to enhance the generalization and reduce overfitting of noisy data [31].
- KNN: a simple example of an instance-based method is the KNN whose name is quite self-explanatory; it classifies a sample with the most common label of the k nearest samples. Though computationally expensive with large text corpora, it is effective in small ones [32].
- DT: a DT is the hierarchical model, which classifies instances by dividing the data recursively based on thresholds of features. It can handle both categorical and numerical data and is easily interpretable [33].
- Linear support vector machine (LSVM): it is a supervised learning algorithm that identifies the optimal hyperplane separating data points of multiple classes in a high dimensional space. This approach is capable of handling both categorical and numerical data and is easily interpretable. It works well on text and other high-dimensional, sparse data [34].
- Kernel support vector machine (KSVM): Kernel SVM, which enhances the basic SVM [35], works at non-linearly inseparable data points by non-linearly mapping them into a higher-dimensional space with a linear separation available using kernel functions.

2.5. Performance evaluation

F1-score, accuracy, precision, and recall were the main metrics used to assess the model's performance. Particular emphasis was placed at first on recall to ensure the model identified most abusive comments correctly and on F1-score to balance precision and recall for accurate classification. Furthermore, we implemented a voting algorithm for finding the best robust model, which was MNB. Due to ensemble results being divided and fairly compared with single models based on unigram, bigram, and trigram feature representations, majority voting across n-grams enhances robustness.

3. RESULT ANALYSIS

Our work achieved a remarkable outcome compared to the previous work in identifying and categorizing abusive comments.

3.1. Performance evaluation metrics

Accuracy, precision, recall, F1 score, and the confusion matrix were among the performance metrics used to evaluate the machine learning model's performance.

- Accuracy: the terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Accuracy measures how often the model correctly predicts the labels.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision: precision refers to the ratio of correctly predicted positive cases to the total predicted positive cases.

$$\text{Precision} = \frac{TP}{TP + FN} \quad (2)$$

- Recall: evaluates how many actual positives were correctly detected.

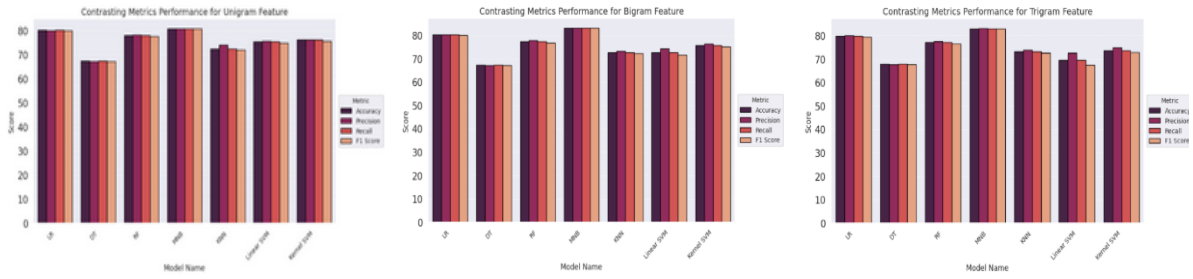
$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

- F1-score: it is a harmonic mean of precision and recall, balancing them.

$$\text{F1-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

3.2. Comparison of performance across various metrics and models

The unigram, bigram, and trigram feature sets were used to test the machine learning models employed in this investigation. The evaluation of their performance was based on accuracy, precision, recall, and F1 score to measure the effectiveness of comment classification. Figure 8 illustrates a visual comparison of these metrics across the three feature representations, where Figures 8(a)-(c) correspond to the unigram, bigram, and trigram feature sets, respectively.



(a) (b) (c)

Figure 8. Comparison of various metrics for (a) unigram, (b) bigram, and (c) trigram

Table 3 shows the analysis and contrasts the outcomes. From the table, performance table of different models after training, the MNB model outperforms with an accuracy of 84.74% for trigram feature extraction, the F1-score stands at 84.66%, and recall at 84.74%. These findings show a strong balance

between minimizing false predictions and accurately detecting abusive content. LR shows competitive results, especially when using bigram features with 80.38% accuracy, but it is still worse than MNB in all the evaluation metrics. Other models show lower performance with significantly worse performance observed when higher order n-gram features are used. Due to its effective sparse handling and high-dimensional TF features, MNB demonstrates high performance. These features render it very suitable to noisy and informal Bangla social media text where spelling differences and colloquialisms prevail.

Table 3. Performance evaluation of various models for unigram, bigram, and trigram features

Model	Unigram				Bigram				Trigram			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
LR	80.23	80.00	80.23	79.99	80.38	80.38	80.38	80.03	79.94	80.02	79.94	79.52
DT	67.42	67.22	67.42	67.17	67.42	67.23	67.42	67.23	67.91	67.65	67.91	67.66
RF	77.98	78.16	77.98	77.65	77.35	77.80	77.35	76.83	77.10	77.66	77.10	76.54
MNB	81.22	81.16	81.22	81.19	84.45	84.43	84.46	84.39	84.74	84.88	84.74	84.66
KNN	72.41	74.08	72.41	72.09	72.70	73.42	72.70	72.31	73.24	73.98	73.24	72.72
LSVM	75.49	75.65	75.49	74.80	72.80	74.30	72.80	71.73	69.57	72.58	69.57	67.45
KSVM	76.27	76.33	76.27	75.72	75.78	76.42	75.78	75.11	73.78	74.99	73.78	72.89

A study of incorrect cases has shown that the larger percentage of errors are in those cases that have sarcasm, subtle abuse, slang, or very colloquial expression. Brief remarks that contain little contextual data also prove difficult and in most cases the predictions are ambiguous. These results demonstrate the innate complexity of the abusive language detection in Bengali social media text. The dataset depicts class imbalance that is familiar in sentiment tasks. To address class imbalance remains an important direction for future research by resampling or cost-sensitive learning.

4. DISCUSSION

In this section, we compare the findings of our study on feminist comment detection with those of prior research, with the aim of highlighting the improvements achieved by our proposed approach. Figure 9 presents a comparative overview of our work alongside previous studies that have focused on abuse detection in different contexts or languages. This demonstrates that our results are more accurate than those of the previous work. Our approach shows a significant improvement in accuracy compared to previous studies that have focused on either general abuse detection or on feminist issues.

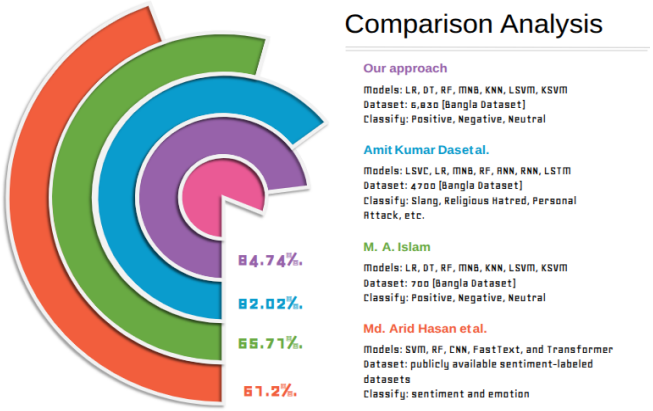


Figure 9. Comparison analysis of previous work

5. CONCLUSION

In this paper, a machine learning model was developed to identify public comments within feminism-related Bangla text, aiming to address a significant challenge in the digital landscape. The research proposed by concentrating on a language underrepresentation and a domain-specific discourse can illustrate that lightweight probabilistic models can be competitive in multi-class language detection in low-resource NLP scenarios. The MNB classifier was found to perform the most successfully with 84.74% accuracy,

which is probably because it works well at modelling sparse TF-IDF features and is robust under limited, high-dimensional text conditions. As much as the model was shown to be good, it mainly uses topical aspects of the lexical and manually marked data, thus being prone to being ineffective in capturing implicit abuse, sarcasm, and contextual nuances. Practically, the experimental performance demonstrates the relevance of the suggested solution to the actual moderation systems on social media in Bengali. However, although this experiment shows promising results, there is still a question. The ability of models for identifying deeper semantic and contextual data is limited by the average size of the dataset and its dependency on TF–IDF features. More research in the future will look into domain-adaptive transformer models and contextual embedding methods to further understand cases of implicit abuse, sarcasm, and ideological dismissal that are frequently found in feminist discourse. Also, sociolinguistic and pragmatic elements, unique to the feminist discourse of Bengali, can be added to the models, which can further increase the robustness of the model.

ACKNOWLEDGMENTS

The authors would like to thank the individuals who supported the annotation process and provided feedback during data validation. Their contributions were invaluable in ensuring the quality of the dataset.

FUNDING INFORMATION

The authors state that there was no funding for the development of the research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Zaid Bin Sajid	✓			✓				✓		✓			✓	
Md. Mijanur Rahman	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	
Md. Sumon Hosen		✓		✓					✓		✓			
Sarara Jaman Riya		✓	✓			✓	✓			✓				
Yeamin Akon		✓	✓			✓	✓			✓				
S. M. Fahad Bin Jim			✓			✓	✓			✓				
Ornab Biswass		✓	✓				✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no financial, personal, or professional conflicts of interest that could have influenced the results or interpretations of this research.

DATA AVAILABILITY

The data that support the findings of this study are openly available in data at https://data.mendeley.com/datasets/nxnbwwc7bn/4.

REFERENCES

[1] A. Akhter, U. K. Acharjee, M. A. Talukder, M. M. Islam, and M. A. Uddin, "A robust hybrid machine learning model for Bengali cyber bullying detection in social media," *Natural Language Processing Journal*, vol. 4, p. 100027, Sep. 2023, doi: 10.1016/j.nlp.2023.100027.

[2] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A deep learning approach to detect abusive Bengali text," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, IEEE, Jun. 2019, pp. 1–5. doi: 10.1109/ICSCC.2019.8843606.

[3] M. Das, S. Banerjee, P. Saha, and A. Mukherjee, "Hate speech and offensive language detection in Bengali," in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2022, pp. 286–296. doi: 10.18653/v1/2022.aacl-main.23.

[4] M. R. Karim, S. K. Dey, T. Islam, M. Shajalal, and B. R. Chakravarthi, "Multimodal hate speech detection from Bengali memes and texts," in *Communications in Computer and Information Science*, A. K. M, B. R. Charavarthi, B. B. C. O'riordan, H. Murthy, T. Urairaj, and T. Mandl, Eds., 2023, pp. 293–308. doi: 10.1007/978-3-031-33231-9_21.

[5] M. T. Akter, M. Begum, and R. Mustafa, "Bengali sentiment analysis of e-commerce product reviews using K-nearest neighbors," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, IEEE, Feb. 2021, pp. 40–44. doi: 10.1109/ICICT4SD50815.2021.9396910.

[6] P. Chakraborty, F. Nawar, and H. A. Chowdhury, "A ternary sentiment classification of Bangla text data using support vector machine and random forest classifier," in *Topical Drifts in Intelligent Computing. ICCTA 2021. Lecture Notes in Networks and Systems*, J. K. Mandal, P. Hsiung, and R. Sankar Dhar, Eds., Singapore: Springer, 2022, pp. 69–77. doi: 10.1007/978-981-19-0745-6_8.

[7] S. Mondal, M. S. Alom, M. M. Alum, and K. L. S. Sunjida, "Multiclass detection of misogynistic Bangla text from YouTube using transformer-based models," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ECCE64574.2025.11014024.

[8] M. S. Islam *et al.*, "Leveraging BanglaBERT: a transformer-based approach for multiclass hate speech detection in Bangla social media," in *2025 4th International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, IEEE, Jun. 2025, pp. 487–492. doi: 10.1109/ACCESS65134.2025.11135879.

[9] A. Islam, M. R. Hossen, M. M. Arif, A. Al Noman, and M. A. Rahman, "BanglaMM-Disaster: a multimodal transformer-based deep learning framework for multiclass disaster classification in Bangla," *arXiv preprint*, p. 2511.21364, 2025.

[10] A. Al Maruf *et al.*, "Hate speech detection in the Bengali language: a comprehensive survey," *Journal of Big Data*, vol. 11, no. 1, p. 97, Jul. 2024, doi: 10.1186/s40537-024-00956-z.

[11] M. A. Hasan, J. Tajrin, S. A. Chowdhury, and F. Alam, "Sentiment classification in Bangla textual content: a comparative study," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2020, pp. 1–6, doi: 10.1109/ICCIT51783.2020.9392681.

[12] V. Shah, A. Sinha, N. Navalkar, S. Gupta, P. Gonsalves, and A. Malik, "ML and natural language processing: cyberbullying detection system for safer and culturally adaptive digital communities," *Journal of Smart Internet of Things*, vol. 2023, no. 2, pp. 193–205, 2023, doi: 10.2478/jsiot-2023-0020.

[13] T. Mahmud, M. Ptaszynski, and F. Masui, "Exhaustive study into machine learning and deep learning methods for multilingual cyberbullying detection in Bangla and Chittagonian texts," *Electronics*, vol. 13, no. 9, p. 1677, Apr. 2024, doi: 10.3390/electronics13091677.

[14] N. R. Bhowmik, M. Arifuzzaman, and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms," *Array*, vol. 13, p. 100123, Mar. 2022, doi: 10.1016/j.array.2021.100123.

[15] WorldData.info, "Bengali speaking countries." Accessed: Jun. 30, 2025. [Online]. Available: https://www.worlddata.info/languages/bengali.php

[16] M. A. Islam, A. C. Aka, A. Akter, and M. M. Rahman, "Analyzing abusive comments in Bangla: machine learning study of feminism on social media," in *Advanced Computing and Intelligent Technologies. ICACIT 2023. Lecture Notes in Networks and Systems*, R. N. Shaw, S. Das, M. Paprzycki, A. Ghosh, and M. Bianchini, Eds., Singapore: Springer, 2024, pp. 379–396, doi: 10.1007/978-981-97-1961-7_25.

[17] A. K. Das, A. Al Asif, A. Paul, and M. N. Hossain, "Bangla hate speech detection on social media using attention-based recurrent neural network," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 578–591, Apr. 2021, doi: 10.1515/jisys-2020-0060.

[18] A. Dhar, H. Mukherjee, R. Bose, S. Roy, and K. Roy, "A study on sentiment analysis using text summarization," in *Proceedings of the 2nd International Conference on Emerging Technologies in Computing (ICETC 2022)*, 2022.

[19] M. M. Rahman, M. S. Hosen, Z. Bin Sajid, and F. F. Sifat, "Social media discourse on feminism: a dataset for sentiment analysis in Bangla comments," *Data in Brief*, vol. 64, p. 112383, Feb. 2026, doi: 10.1016/j.dib.2025.112383.

[20] G. B. Herwanto, A. M. Ningtyas, K. E. Nugraha, and I. N. P. Trisna, "Hate speech and abusive language classification using fastText," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2019, pp. 69–72. doi: 10.1109/ISRITI48646.2019.9034560.

[21] O. Faruqe, M. Jahan, M. Faisal, M. S. Islam, and R. Khan, "Bangla hate speech detection system using transformer-based NLP and deep learning techniques," in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, IEEE, Aug. 2023, pp. 1–6, doi: 10.1109/ASIANCON58793.2023.10269919.

[22] O. Levchenko and M. Dilai, "Attitudes toward feminism in Ukraine: a sentiment analysis of tweets," in *Advances in Intelligent Systems and Computing*, Cham: Springer, 2019, pp. 119–131, doi: 10.1007/978-3-030-01069-0_9.

[23] W. J. Scarborough, "Feminist Twitter and gender attitudes: opportunities and limitations to using Twitter in the study of public opinion," *Socius: Sociological Research for a Dynamic World*, vol. 4, Jan. 2018, doi: 10.1177/2378023118780760.

[24] N. Albadi, M. Kurdi, and S. Mishra, "Are they our brothers? Analysis and detection of religious hate speech in the Arabic Twittersphere," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, IEEE, Aug. 2018, pp. 69–76. doi: 10.1109/ASONAM.2018.8508247.

[25] M. Hassan *et al.*, "Sentiment analysis on Bangla conversation using machine learning approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 5562–5572, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5562-5572.

[26] A. N. Tusher, S. Islam, M. T. Islam, M. S. R. Sammy, M. S. Rahman, and M. S. Sadik, "User perspective Bangla sentiment analysis for online gaming addiction using machine learning," in *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, IEEE, Nov. 2022, pp. 538–543, doi: 10.1109/I-SMAC55078.2022.9987343.

[27] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter, and A. K. Das, "An automated system of sentiment analysis from Bangla text using supervised learning techniques," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, IEEE, Feb. 2019, pp. 360–364. doi: 10.1109/CCOMS.2019.8821658.

[28] Analytics Vidhya, "Unlock free ensemble learning algorithm course." Accessed: Jul. 02, 2025. [Online]. Available: https://courses.analyticsvidhya.com/courses/ensemble-learning-and-ensemble-learning-techniques

[29] L. Zhang, "Features extraction based on Naive Bayes algorithm and TF-IDF for news classification," *PLOS One*, vol. 20, no. 7, p. e0327347, Jul. 2025, doi: 10.1371/journal.pone.0327347.

[30] N. Amin, "Comparative evaluation of logistic regression and Naïve Bayes for fake news detection using NLP techniques," *Preprints*, p. 2025120630, Dec. 2025, doi: 10.20944/preprints202512.0630.v1.

[31] H. Allam, L. Makubvure, B. Gyamfi, K. N. Graham, and K. Akinwolere, "Text classification: how machine learning is revolutionizing text categorization," *Information*, vol. 16, no. 2, p. 130, Feb. 2025, doi: 10.3390/info16020130.
[32] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, Aug. 2024, doi: 10.1186/s40537-024-00973-y.
[33] S. K. S. Joy *et al.*, "Gender bias mitigation for Bangla classification tasks," *arXiv preprint*, p. 2411.10636, 2024.
[34] S. Fatima and B. Srinivasu, "Text document categorization using support vector machine," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 2, pp. 141–147, 2017.
[35] F. Maggioni and A. Spinelli, "A novel robust optimization model for nonlinear support vector machine," *European Journal of Operational Research*, vol. 322, no. 1, pp. 237–253, Apr. 2025, doi: 10.1016/j.ejor.2024.12.014.

BIOGRAPHIES OF AUTHORS

Zaid Bin Sajid completed his bachelor's degree in computer science and engineering from Southeast University. He has been working as a graduate research assistant at Southeast University in Dhaka, Bangladesh. His interests include networking, IoT, machine learning, deep learning, NLP, and blockchain. He has published research in the field of blockchain and is currently working on multiple papers involving machine learning and deep learning. He can be contacted at email: zaidbinsajid1@gmail.com.

Md. Mijanur Rahman is currently an assistant professor in the Department of Computer Science and Engineering and Director of the Center for Artificial Intelligence and Robotics at Southeast University. His research interests include deep learning, machine learning, medical image processing, data science, recommendation systems, and blockchain. Rahman has published more than 45 peer-reviewed articles in conferences and journals. He can be contacted at email: mijanur.rahman@seu.edu.bd.

Md. Sumon Hosen completed his bachelor's degree in computer science and engineering from Southeast University. He has been working as a graduate research assistant at Southeast University in Dhaka, Bangladesh. His interests include machine learning, deep learning, data science, and NLP. He has also contributed to research through his recent publication in the field. He can be contacted at email: csesumon8@gmail.com.

Sarara Jaman Riya holds a bachelor's degree in computer science and engineering from Southeast University. She is interested in machine learning and deep learning and actively follows the latest advancements in these areas. She can be contacted at email: riyasarara@gmail.com.

Yeamin Akon holds a bachelor's degree in computer science and engineering from Southeast University. He is interested in data science, machine learning, and deep learning, and stays engaged with emerging developments in these fields. He can be contacted at email: yaffianafyeamin@gmail.com.

S. M. Fahad Bin Jim holds a bachelor's degree in computer science and engineering from Southeast University. His interests include data science, machine learning, and deep learning. He can be contacted at email: fahadbinjim9@gmail.com.

Ornab Biswass holds a bachelor's degree in computer science and engineering from Southeast University. He is interested in data science, machine learning, and deep learning, and stays engaged with emerging technological advancements in these fields. He can be contacted at email: 2020000000095@seu.edu.bd.