Computing and Informatics, Vol. 42, 2023, 993–1012, doi: 10.31577/cai_2023_4_993

DEEP LEARNING BASED MISOGYNISTIC BANGLA TEXT IDENTIFICATION FROM SOCIAL MEDIA

Sarif Sultan Saruar Jahan, Raqeebir Rab, Peom Dutta,
Hossain Muhammad Mahdi Hassan Khan,
Muhammad Shahariar Karim Badhon

Ahsanullah University of Science and Technology
Department of Computer Science and Engineering
Dhaka, Bangladesh
e-mail: sjalim71@gmail.com, raqeebir.cse@aust.edu, {peomd04, mahdihassankhan10, sms.badhon}@gmail.com

Sumaiya Binte Hassan

The University of British Columbia
UBC Brain, Attention, and Reality Lab
Vancouver, Canada
e-mail: sumaiya.b.hassan@gmail.com

Ashikur Rahman

Bangladesh University of Engineering and Technology
Department of Computer Science and Engineering
Dhaka, Bangladesh
e-mail: ashikur@cse.buet.ac.bd

Abstract. Misogyny is characterized by hostility, hatred, aversion, intimidation, and violence against women. With the rise of social media, it has become one of the most convenient platforms for expressing woman-hating speech. As a result, misogyny is gaining appeal and societal standards are being violated. With millions of Bangladeshi Facebook users, misogyny is growing increasingly prevalent in Bangla

as well. In this paper, we have proposed automatically identifying misogynistic content in Bangla on social media platforms in order to evaluate the problem's challenges. As there is no existing Bangla dataset for analyzing misogynistic text, we generated our own. We have applied various deep-learning algorithms to improve the classification of misogynistic text categories. LSTM and RNN models are used for designing the model architecture in deep learning. Models are evaluated using the confusion matrix, accuracy, and f1-scores. The results indicate that LSTM outperforms RNN in terms of accuracy by 67 %.

Keywords: Misogyny, deep learning, LSTM, RNN, BERT, feature selection, natural language processing

1 INTRODUCTION

Misogyny is the hate or prejudice against women that can be linguistically manifested in numerous ways, ranging from less aggressive behaviors like social exclusion and discrimination to more dangerous expressions related to threats of violence and sexual objectification. In recent years, misogyny has increased exponentially due to the widespread global use of social media platforms such as Twitter, Facebook, Instagram, and YouTube. Misogyny appears in different forms in our society, causing incalculable harm to girls and women. In this case, social media may have been a method to guarantee free speech but social media sites must now monitor and prohibit abusive content to safeguard their users. The proliferation of misogynistic content online causes an increase in social misbehavior, promoting and instigating actual hate crimes. It creates an association between the rise of misogynistic conduct online and the number of rapes in the United States [1]. Currently, the detection of women-targeted harassment in social networks is receiving increasing attention.

Misogynistic text classification is a widely researched topic, with the majority of research conducted in European languages such as English, French, Italian, etc. It is rarely studied in the context of the Bengali language, despite the fact that it is widely spoken. Bangladesh has a population of 168.7 million, and 31.5 % of the population utilizes the internet, which translates to around 58.7 million internet users by 2022 [2]. The annual growth rate of active social media users is 25 percent (9 million) [3]. As social media platforms grow in size, misogynistic behavior is increasingly reflected in the Bangla language on these sites. In Bangladesh, almost 75 % of females use social media multiple times daily, compared to 64 % of males which is a substantial proportion [4]. While new opportunities for women's self-expression have emerged, misogyny manifests as a categorization of the feminine gender. 73 % of female internet users are victims of various cyber crimes [5]. 30 % are unfamiliar with the methods for getting assistance [3]. The psychological effects of online hateful speech extend beyond the victims to the readers as well. Due to

these forms of social media activities, negative affect disorders, loneliness, anxiety, depression, suicidal ideation, and somatic symptoms are prevalent among female social media users. Misogynistic texts can have devastating societal and personal consequences. These manifestations of misogyny are a pertinent social issue that has been explored in the scientific literature during the past several years.

In this paper, we identified the label of a text based on social media comments and classified it into one of the three categories: Stereotype and Objectification, Dominance, Derailing, Sexual Harassment, and Discredit.

We have proposed deep learning-based methods for building a multi-class analyzer, where deep Learning algorithms such as RNN, LSTM, and BERT word embeddings are employed to identify misogynistic text. In recent years, deep learning techniques have performed exceptionally well because they do not require predefined characteristics; rather, they acquire knowledge from the dataset itself. We analyze the performance of our technique on a dataset containing Bengali-language comments from several social media platforms. After thorough experiments, the RNN model exhibits 55 % accuracy and LSTM with 67 % accuracy. The following is a summary of the primary contributions of this work:

1. Construct a dataset of misogynistic text in the Bangla language for the first time.
2. Approach the automatic detection of misogyny against women using a deep-learning approach; and
3. Evaluate the performance of our technique on a dataset consisting of Bengali-language comments from multiple social media sites.

The remaining paper is organized as follows: Section 2 is a summary of the relevant work. The dataset is described in Section 3. Section 4 discusses the methodology for detecting misogyny against women. Section 5 includes the experiments conducted and analyzes the results collected. Section 6 provides a discussion of the analysis and its conclusions.

2 RELATED WORKS

There is not much work that exists on misogynistic text in the Bengali language. In contrast, English, Italian, and other languages have a substantial amount of work in this field. The majority of work is concentrated on the dataset based on Twitter. For purposes of detection, machine learning techniques such as Logistic Regression, Support Vector Machine (SVM), and Naive Bayes classifier have dominated. These classifiers are trained alongside the TF-IDF word vectorization technique and integrated with the models [2]. In addition, other approaches utilized Bag and sequences of words, Characters n-grams, and Lexicons for classification. [6] utilized three data sets collected from Twitter. The SVM Model has been implemented in TF-IDF. The data has been preprocessed with Natural Language

Toolkit (NLTK)[1]. Information Gain SVM was used to calculate the weights of lexical characteristics in order to detect misogynistic tweets. Bakarov [7] approach is designed to identify misogynistic text collected from Twitter. Their system is based on a vector space model of character n-grams and a supervised gradient-boosting classifier. Another study by Frenda et al. [8] also involved two subtasks: Misogyny Identification and Misogynistic Behavior and Target Classification to identify misogynistic content on Twitter in English and Italian. They applied lexica modeling to enrich the dictionary and used an SVM classifier with RBF for each language. The paper of Ahluwalia et al. [9] collected 5 000 tweets from Twitter and classified tweets as misogynistic or not using NLTK for tokenization. Feature extraction involved Bag of Words. They applied different Machine Learning, Deep Learning, and Ensemble Learning models for classification. The best result was obtained through Deep Learning. On the other contrary, Alawneh et al. [10] collected 4 000 labeled data from "maps.safecity" categorized into Ogling, Commenting, and Groping. Data pre-processing involved tokenization, stemming, and lemmatization. Tf-Idf was used for feature extraction, and eight classifiers, including Random Forest, Multinomial NB, SVS, Linear SVC, SGD, Bernoulli NB, DT, and K-Neighbors, were evaluated. The proposed model achieved 81 % accuracy using the SGD classifier.

The current study has also focused on Deep Learning approaches for misogynistic text detection. Ordered Neurons LSTM with XLM-RoBERTa was proposed by Ou and Li [11] for hate speech detection using a dataset that included 6 839 tweets. The K-max pooling and convolution neural network was constructed using the pre-trained multilingual model XLMRoBERTa. A linear decision function is applied following the addition of an Ordered Neurons LSTM (ONLSTM) to the prior representation. In the multilingual environment, Datta et al. [12] conducted a detection study in three languages – English, Hindi, and Bangla – using a dataset of 18 000 labeled instances categorized into Overtly aggressive, Covertly aggressive, and not aggressive. Throughout the classification process, different feature models were employed for each language, and Tf-Idf was always utilized for feature extraction. The XGBoost Classifier was utilized for English Text Classification, whereas the Gradient Boosting Classifier (GBC) was employed for Bangla and Hindi Text Classification. On the English dataset, the model achieved 58 % accuracy, on the Hindi dataset, 62.08 % accuracy, and on the Bangla dataset, 59.76 % accuracy. Yet, deeper learning could produce superior results.

Chakraborty and Seddiqui [13] employed Multinomial Naive Bayes (MNB), Support Vector Machine (SVM), and CNN-LSTM as classification algorithms in a balanced dataset of 5644 instances where 50 % were labeled as 'threat and abuse' and the rest as 'No'. The SVM classifier showed the most consistent performance with an accuracy of 78 %, which was the highest achieved among the models used. Fersini et al. [14] performed a classification task to predict Misogyny and Aggressiveness in a dataset of 4 000 samples. Pre-processing was done using Word to Vector and

[1] https://www.nltk.org/

feature extraction using Tf-Idf. Three machine learning models were used, a Shallow Model, a Convolutional Neural Network, and Fine-tuning of the Pre-trained model (a multilingual strategy using BERT). The best results were obtained using the Fine-tuning of the Pre-trained model with BERT.

The majority of relevant NLP research focuses on classifying misogynistic text in tweets written in English and other languages. Misigonystic Bengali text within social media has not been explored much. We analyzed misogynistic Bengali text in social media and developed a data set of misogynistic Bengali comments as a baseline for our research.

3 DATASET

There was no Bengali dataset for identifying misogynistic text. Thereby, we created a whole new dataset for our research. The most well-known social media platforms, Facebook, Instagram, TikTok, and YouTube public posts have been used for collecting the data. We focused on how individuals acted and thought about women, as seen by their comments. We rely heavily on the opinions of individuals associated with misogyny. To differentiate between misogynistic and non-misogynistic texts, we attempted to select non-misogynistic data that was largely linked with women.

3.1 Data Acquisition

We have collected approximately *4 000* raw data from Facebook, Instagram, TikTok and YouTube. The 55 k comments on public posts from these platforms are key source of our data. Initially, we labeled the text manually. To validate these labels we have conducted a survey. According to the survey results, we have amended our dataset to include the previously mislabeled texts. Moreover, a sociologist has validated this dataset. Furthermore, as we intended to detect these texts using deep learning models the size of the dataset was not enough. To overcome this issue we have augmented the dataset using some pre-trained BERT models. Finally, in total the dataset size is *15 k*. Some of the post links can be found here[2]. Figure 1 shows some of survey report.

3.2 Data Cleaning

Data cleaning is a crucial stage in our process. Several comments contain misspellings and a combination of languages, including Bangla, English, and Banglish. All of them were manually corrected. Table 1 shows some examples of data cleaning.

[2] `https://shorturl.at/jwNUZ`

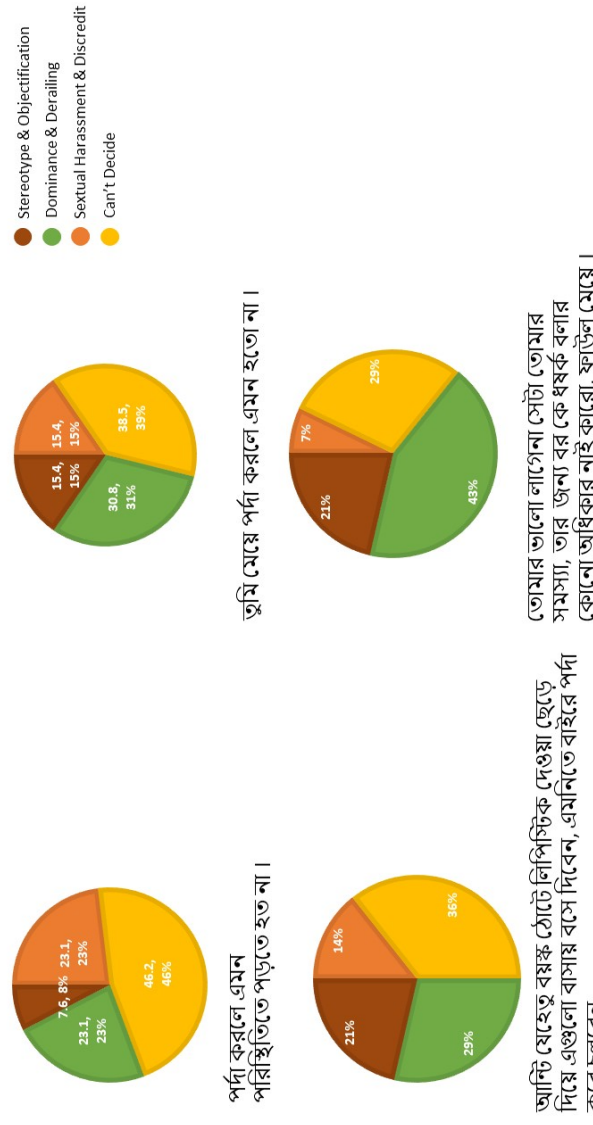

Figure 1. Sample survey report

Raw Text	Processed clean Text
যোবন তমার লাল টমাটু	যৌবন তোমার লাল টমেটো
তুই জানোস rape e মেয়েরা আসলে অনেক মজা পায়	তুই জানোস ধর্ষণে মেয়েরা আসলে অনেক মজা পায়
nari voge er jonno	নারী ভোগের জন্য

Table 1. Preview of clean processed text

3.3 Annotation Process

Depending on the misogynistic text, we grouped the misogynistic texts into three categories: stereotype and objectification, dominance, derailment, sexual harassment and the threat of violence, and discredit. Anzovino et al. [15] defined seven categories of misogynistic texts. The Bengali language contains all seven forms of misogynistic data, but some of the definitions are inconsistent with Bengali text. Hence, we have merged several categories in the Bengali language context. After examining the entire of our dataset's text we have ***three*** target classes for misogynistic texts and one for non-misogynistic texts. These categories are addressed in the next section in perspective of the considerations we have made based on the definitions mentioned by Anzovino et al. [15].

Stereotype and Objectification: Stereotypes consist of preconceived conceptions and unjustified assumptions about women, in addition to the rigid and unsophisticated image or depiction of a woman. Objectification is the act of comparing women's physical characteristics to a limited set of criteria and preconceived assumptions regarding a woman's competence for a particular task.

Example: ঘরের লক্ষীদের ঘরেই শোভা পায়, বাসে না (Female are suitable at home, not in a bus.)

Dominance and Derailing: This category consists of objects that no one has the authority to impose on women, the belief that men are superior to women, and the emphasis on gender inequity. The appearance of women is one of the primary factors upon which people form their opinions. The term dominance refers to any expression of a dominant attitude. Otherwise, we would have considered this content objectifying and stereotypical. Dominance is described as the attempt to alter a woman's words to make them more acceptable to men or the defense of any abusive behavior toward a woman. Derailment is the denial of manly accountability.

Example: তারা বের করে চলবেন আর আমরা দেখলে (They will walk out and we will see the rape.)

Sexual Harassment, Threat of Violence and Discredit: Texts that sexually insult women, solicitations for sexual favors, sexually explicit harassment, and the classification of certain activities as sexual advances are all examples of sexual harassment. The intent to physically dominate women through intimidation

is a violent threat. Women-directed abuses without a larger purpose are considered disrespectful.

Example: সুন্দরী এবং রূপবতী মেয়েরা আসলে মাগি, (Beautiful and beautiful girls are actually witches, whores and harlots.)

3.4 Data Validation

We have followed three steps in validation:

1. Comprehensive recheck of the entire annotated dataset according to the definition mentioned by Anzovino et al. [15].
2. Conducted survey on the dataset with men and women of varying ages. We updated the entire dataset based on the results of the survey.
3. We verified our level data by Sumaiya Binte Hassan, UBC Brain, Attention and Reality Lab, University of British Colombia, Vancouver, Canada.

3.5 Data Augmentation

Our data collection is quite limited for training deep learning models. Deep learning techniques require large amounts of information to enhance accuracy. Finding raw misogynistic text in Bangla is similarly challenging. Hence, one technique to increase the amount of data is to artificially add data to our dataset. Data augmentation is a well-known method for producing synthetic data from an existing dataset. In this case, Python libraries have been used. Pretrained BERT models are utilized for the purpose of augmentation. The Code snippet of data augmentation is provided[3].

Three Bengali pre-trained models were used to augment the text. Those are:

- banglabert [16],
- bangla-bert-base [17],
- sahajBERT[4].

3.6 Data Distribution

As shown in Figures 3 and 4, we separated our dataset into two parts: a binary dataset including misogynistic and non-misogynistic text. Another is a multi-class dataset in which text was categorized according to stereotype and objectification, dominance, derailment, sexual harassment and threat of violence, and discredit. The main corpus distribution is shown in Figure 2.

[3] https://github.com/sjalim/ColabCode/blob/main/text_augmentation.ipynb
[4] https://huggingface.co/neuropark/sahajBERT

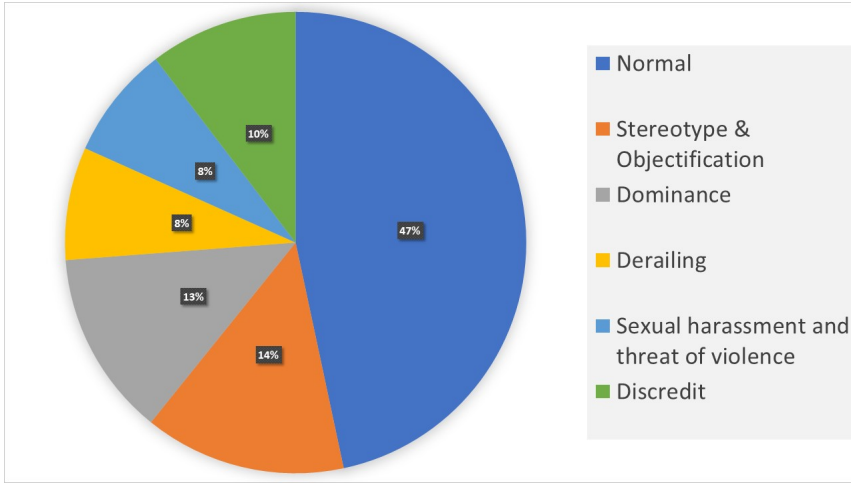

Figure 2. Main corpus distribution

4 METHODOLOGY

The purpose of this research is to identify the three types of misogynistic texts directed against women. We utilized Deep Learning techniques for classification. RNN and LSTM models are employed in the design of the neural network. Whereas the dataset was built from scratch, data pre-processing is required to obtain cleaner data. The BERT embeddings pre-trained model was used for contextual text com-

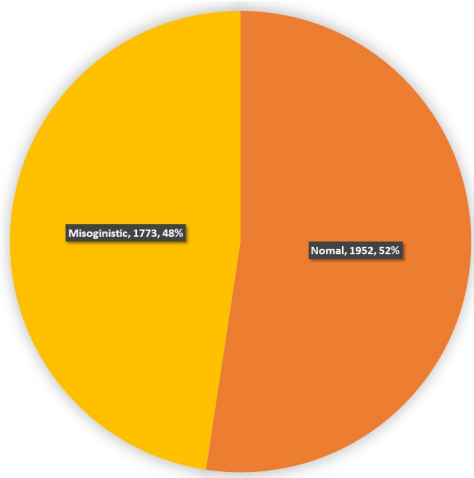

Figure 3. Binary corpus distribution

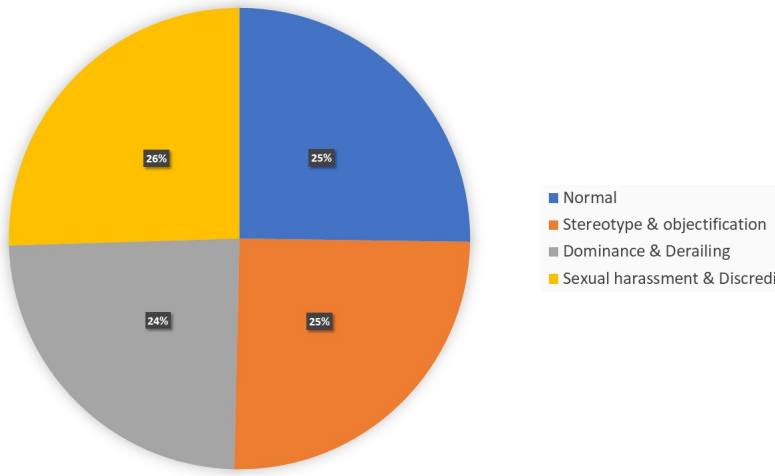

Figure 4. Multi corpus distribution

prehension in the models. Optimizers and activation functions also played a crucial role in terms of fine-tuning. The overview working procedure of our model is depicted in Figure 5.

4.1 Data Pre-Processing

Our information was gathered from several social media sites, thus it contains characters, numbers, and text fragments that can be categorized as noise and interfered with text processing. We have applied a few standard preprocessing approaches to reduce the noise.

4.1.1 POS Tagger

Sentences contain numerous non-dominant parts of speech, including conjunctions, interjections, prepositions, nouns, and pronouns, among others. When the corpus grows, the entire form of text might make text analysis cumbersome. We employed Taggers for Parts of Speech (POS tags) to identify the lexical terms included in text[5].

Examples of different types of lexical terms, their tags from Bangla Text is given below:

- Noun: মিনা, রিনা;
- Interjection: হায়, ওহ, ওঃ, ওমা;
- Pronoun: আমি, তুমি, তোমার, তার;

[5] https://bnlp.readthedocs.io/en/latest/

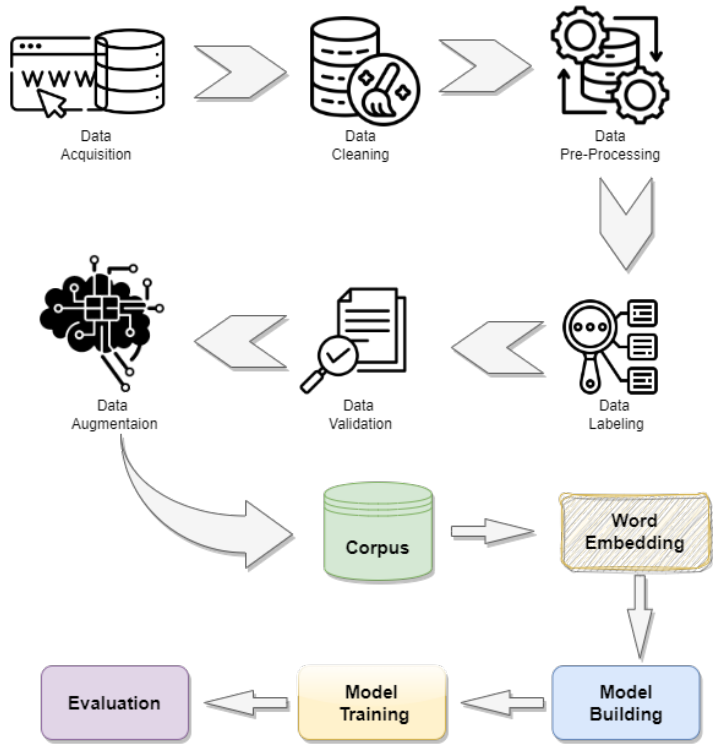

Figure 5. Work-flow diagram for the proposed model

- Preposition: দ্বারা, দিয়ে, তবে;
- Conjunction: ও, আর, এবং.

4.1.2 Removing Numbers, Punctuation, and Emoji

Numbers, punctuation, and emoji for example ১৩, 54, : ;, ☺, 🚀 are irrelevant to our work. They have no contextual significance in texts that are misogynistic. Usually, they are regarded as noise in text. The removal of numerals, punctuation, and emoji improves the performance of our model.

4.2 Word Embeddings

To extract the semantic alignment of word vectors, we have employed word embedding. For word embedding, BERT-trained models were utilized. There are numerous available strategies for feature extraction. Nonetheless, word embedding is a highly effective method for text comprehension. We have word vectorization techniques such as TF-IDF, BOW, etc. based on the frequency of words. The context of a sen-

tence is lost when its frequency is measured. On the other hand, BERT extracted contextualized word embeddings via transfer learning [18]. We have transformed every text using sbert[6] sentence transformer [19].

On our dataset, three pre-trained BERT embedding models such as BanglaBERT, Bangla BERT Base, and SahajBERT were evaluated for generating word embeddings in Bengali text. These three models performed the best with the proposed model architecture for deep learning.

4.2.1 BanglaBERT

This model gives 768 embeddings for every sentence. They have trained 27.5 GB bangla data from different websites [16].

4.2.2 Bangla Bert Base

This model gives 768 embeddings for every sentence. They have used Wikipedia Dump Dataset and OSCAR[7] Bengali dataset to train this pre-trained model [17].

4.2.3 SahajBERT

This model gives 1024 embeddings for every sentence. Collaboratively pre-trained model[8] on Bengali language using masked language modeling (MLM) and Sentence Order Prediction (SOP) objectives. They have also used Wikipedia and OSCAR datasets.

4.3 Model Architecture

Since we work with texts, sequence ordering is crucial in this case. We designed the neural network architecture for text classification using RNN and LSTM models.

RNN can learn from sequential data. In the core of its architecture, however, we frequently observe hidden layers in which entire learned knowledge is transferred from one layer to another. This can lead to issues. Because one layer may not require the knowledge of another layer to comprehend the intricate context of a sentence. In contrast, if part of the layers did not learn anything from the text, they would send a gradient of zero to the following layer, resulting in a problem with vanishing gradients. So initially, we utilized the RNN layer for model development. We have employed the LSTM model to further improve this model.

LSTM model with cell state resolved the issue. It can calculate the quantity of information that must be sent together with the cell's status. Other than this, the

[6] https://www.sbert.net/

[7] https://oscar-project.org/

[8] https://huggingface.co/neuropark/sahajBERT

top-level architecture is comparable to the RNN model. In addition to the RNN and LSTM models, our design consists of five (5) linear layers. Using ReLU and Softmax activation routines. Our model architecture is shown in Figure 6.

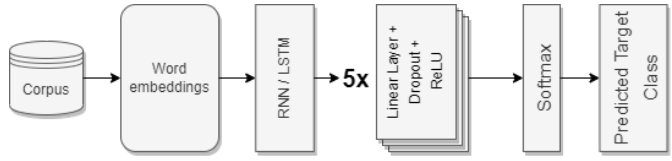

Figure 6. Model architecture

In addition to the architecture, we have experimented with various optimizers, including Adam, RAdam, and NAdam. Despite the fact that we are dealing with a multi-class issue we utilized Cross Entropy Loss for loss computation. As PyTorch can handle sparse target labels, it was utilized to implement the entire model architecture. We used a learning rate of 0.001, a batch size of 10, and 10 epochs to train the model. Around one million parameters were included in the model. Table 2 shows RNN Uni-Directional and RNN Bi-directional models parameter values. We have Table 3 for LSTM Uni-directional and LSTM Bi-directional. Finally, we have Table 4 for parameter values.

Layer	**Text Size**	**Weight Matrix Dimensions**	**Bias Vector Dimensions**
Embedding	Vocab Size×Input Size	Vocab Size×Input Size	Input Size
RNN	Hidden Size	Input Size × Hidden Size	Number of Layers × Hidden Size
Fully Connected 1	512	Hidden Size × 512	512
Fully Connected 2	256	512 × 256	256
Fully Connected 3	128	256 × 128	128
Fully Connected 4	64	128 × 64	64
Fully Connected 5	Number of Classes	64×Number of Classes	Number of Classes

Table 2. RNN parameters

5 EXPERIMENTAL RESULTS AND DISCUSSION

Our research is carried out using RNN and LSTM deep learning techniques for three different pre-trained BERT word embedding models. Accuracy, F1 score, and Cohen Kappa were used to analyze our findings. Our dataset has been divided into training

Layer	**Text Size**	**Weight Matrix Dimensions**	**Bias Vector Dimensions**
Embedding	Vocab Size x Input Size	Vocab Size × Input Size	Input Size
LSTM	Hidden Size	(Input Size + Hidden Size) × 4 × Hidden Size	4 × Number of Layers × Hidden Size
Fully Connected 1	512	Hidden Size × 512	512
Fully Connected 2	128	512 × 128	128
Fully Connected 3	Number of Classes	128 × Number of Classes	Number of Classes

Table 3. LSTM parameters

Parameter Name	**Value**
Vocab Size	15 000
Hidden Size	768
Input Size	768
Number of Layers	2
Number of Classes	2 (Binary)/4(Multi-Class)

Table 4. Parameters value

and testing portions as indicated in Table 5. 20 % of our corpus is used for testing, while 80 % is used for training.

We have conducted several experiments with our corpus. We followed some of the key steps such as a) Model Parameter tuning, b) Dataset reforming, c) Analysis results, etc. Regarding these steps' outcomes, we have made necessary changes to our model as well as to the corpus. The best possible result using our models has been portrayed by Table 6 and Table 7. Here, we see Table 6 is for Binary target class corpus and Table 7 is Multi-class.

Purpose	Count (Binary Class)	Count (Multi-Class)
Training Data	11 894	7 638
Testing Data	2 974	1 910

Table 5. Data distribution

From all the experiments it can be said that our corpus is performing well with the LSTM model where as the RNN model is lacking a bit comparatively. But if we see depending on different pre-trained BERT models the output differs. Those pre-trained models were trained with the different datasets we have mentioned in the above dataset creation section.

We have observed that all three of the BERT pre-trained model **BanglaBert-Base** performed very well. To evaluate our model's result several evaluation matrices

have been used such as Confusion Matrix, Accuracy, F1-Score (macro, weighted), and Cohen Kappa score.

Confusion Matrix gives us a proper understanding of model predicted result. Here, in Figure 7 matrix diagonally we see greater values compared to the second diagonal. A few data were unable to predict by our models. Though we have used the same corpus to train, the results are better with the LSTM model.

On the other side, we have Figure 8 matrix that shows the multi-class performance by our model with our corpus. Similarly, we see the diagonally metric is far better than other cells. The LSTM model performed better than the RNN model. Here the pre-trained model used in both is the same, but, depending on the model, the result defers.

$$\kappa = (p_o - p_e)/(1 - p_o). \tag{1}$$

Three optimizers were used, and among those, **Adam** shows huge potential. All the experiments using Adam perform phenomenally.

Cohen Kappa evaluation metric is used which is a subtle version of accuracy. Equation (1) is the formula to calculate the result, where p_e is the predicted agreement when both annotators issue labels randomly, and p_o is the empirical probability of agreement on the label assigned to each sample calculated using an empirical per-annotator prior over the class labels. Micro, weighted F1-Score has been used to evaluate the model. As we are working with multi-class classification problems macro and weighted value gives a better understanding of classwise performance.

Pre-trained	Model + Optimizer	Accuracy	F1-Score	Cohen Kappa
BanglaBert	RNN + Radam	77.17 %	73.11 %	53.59 %
	LSTM + Nadam	77.37 %	74.88 %	54.34 %
BanglaBertBase	RNN + Radam	77.27 %	74.06 %	53.97 %
	LSTM + Adam	**82.59 %**	**81.17 %**	**64.98 %**
SahajBert	RNN + Adam	78.21 %	76.15 %	56.11 %
	LSTM + Nadam	79.62 %	76.77 %	58.75 %

Table 6. Result of Binary class target

Pre-trained	Model + Optimizer	Accuracy	F1-Score (Macro)	F1-Score (Weighted)	Cohen Kappa
BanglaBert	RNN + NAdam	55.28 %	54.34 %	54.82 %	40.27 %
	LSTM + Adam	54.71 %	53.90 %	54.22 %	38.78 %
BanglaBertBase	RNN + RAdam	55.23 %	54.97 %	55.76 %	40.28 %
	LSTM + Adam	**67.27 %**	**66.91 %**	**67.11 %**	**56.26 %**
SahajBert	RNN + Adam	54.97 %	54.26 %	54.80 %	39.78 %
	LSTM + NAdam	56.38 %	55.82 %	55.74 %	41.76 %

Table 7. Result of Multi-class target

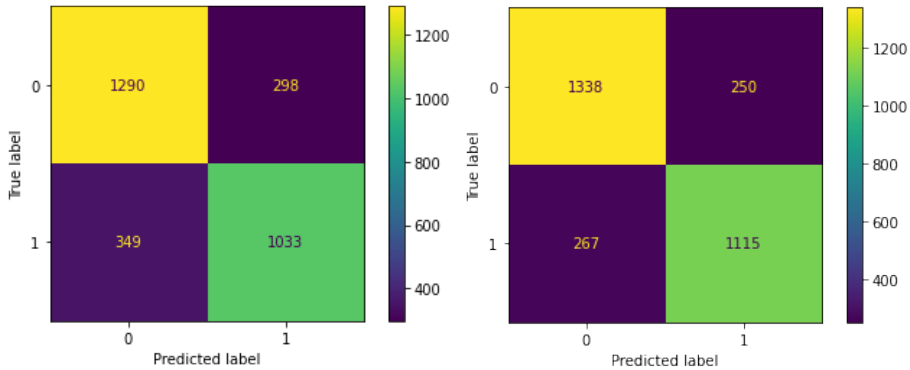

Figure 7. Left: SahajBert (RNN), Right: BanglaBertBase (LSTM) Confusion Matrix

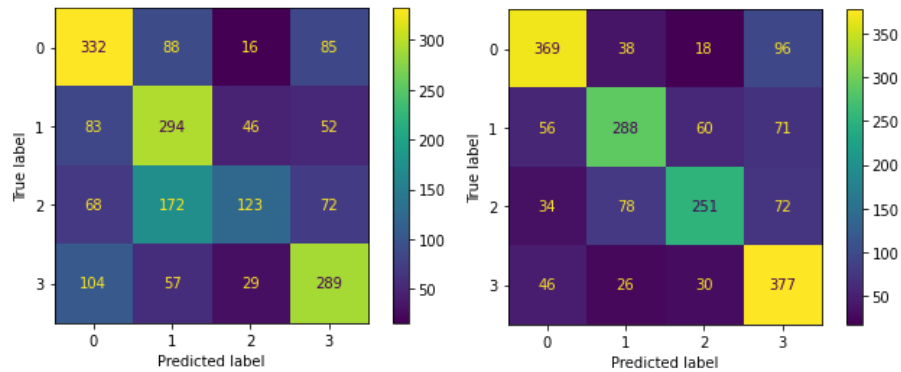

Figure 8. Left: BanglaBertBase (RNN), Right: BanglaBertBase (LSTM) Confusion Matrix

6 CONCLUSION AND FUTURE WORK

We have represented the dataset creation process on misogynistic text as well as proposed deep learning model architecture with LSTM and RNN models. Binary class and Multi-class targets both have been portrayed in this paper. The binary class target accomplished the highest accuracy of 90 % with the LSTM model. On the other hand, multi-class targets accomplished 94 % accuracy with the LSTM model. We have shown the performance of pre-trained BERT models with our dataset. As our dataset size is not enough for deep learning work at its best, we intend to increase the dataset. Thus our model is going to perform much better than now. Also, we will build a Google Chrome add-on to integrate our Misogynistic Text detection system and deploy the system as soon as possible.

REFERENCES

[1] Fulper, R.—Ciampaglia, G. L.—Ferrara, E.—Ahn, Y. Y.—Flammini, A.—Menczer, F.—Lewis, B.—Rowe, K.: Misogynistic Language on Twitter and Sexual Violence. Proceedings of the ACM Web Science Workshop on Computational Approaches to Social Modeling (ChASM'14), 2014.

[2] Shushkevich, E.—Cardiff, J.: Misogyny Detection and Classification in English Tweets: The Experience of the ITT Team. In: Caselli, T., Novielli, N., Patti, V., Rosso, P. (Eds.): EVALITA Evaluation of NLP and Speech Tools for Italian. Proceedings of the Final Workshop (EVALITA 2018). Accademia University Press, Torino, Italy, 2018, pp. 182–187, doi: 10.4000/books.aaccademia.4670.

[3] Akter, F.: Cyber Violence Against Women: The Case of Bangladesh. 2018, `https://genderit.org/articles/cyber-violence-against-women-case-bangladesh`.

[4] Auxier, B.—Anderson, M.: Social Media Use in 2021. 2021, `https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/`.

[5] Halim, T.: 73% Women Face Cyber Crimes: Tarana. 2017, `https://www.thedailystar.net/country/73-women-face-cyber-crimes-tarana-1372849`.

[6] Frenda, S.—Ghanem, B.—Montes-y Gómez, M.—Rosso, P.: Online Hate Speech Against Women: Automatic Identification of Misogyny and Sexism on Twitter. Journal of Intelligent and Fuzzy Systems, Vol. 36, 2019, No. 5, pp. 4743–4752, doi: 10.3233/JIFS-179023.

[7] Bakarov, A.: Vector Space Models for Automatic Misogyny Identification. In: Caselli, T., Novielli, N., Patti, V., Rosso, P. (Eds.): EVALITA Evaluation of NLP and Speech Tools for Italian. Proceedings of the Final Workshop (EVALITA 2018). Accademia University Press, Torino, Italy, 2018, pp. 211–213, doi: 10.4000/books.aaccademia.4740.

[8] Frenda, S.—Ghanem, B.—Guzmán-Falcón, E.—Montes-y Gómez, M.—Villaseñor-Pineda, L.: Automatic Expansion of Lexicons for Multilingual Misogyny Detection. In: Caselli, T., Novielli, N., Patti, V., Rosso, P. (Eds.): EVALITA Evaluation of NLP and Speech Tools for Italian. Proceedings of the Final Workshop (EVALITA 2018). Accademia University Press, Torino, Italy, 2018, pp. 188–193, doi: 10.4000/books.aaccademia.4680.

[9] Ahluwalia, R.—Soni, H.—Callow, E.—Nascimento, A.—De Cock, M.: Detecting Hate Speech Against Women in English Tweets. In: Caselli, T., Novielli, N., Patti, V., Rosso, P. (Eds.): EVALITA Evaluation of NLP and Speech Tools for Italian. Proceedings of the Final Workshop (EVALITA 2018). Accademia University Press, 2018, doi: 10.4000/books.aaccademia.4698.

[10] Alawneh, E.—Al-Fawa'reh, M.—Jafar, M. T.—Al Fayoumi, M.: Sentiment Analysis-Based Sexual Harassment Detection Using Machine Learning Techniques. 2021 International Symposium on Electronics and Smart Devices (ISESD), IEEE, 2021, pp. 1–6, doi: 10.1109/ISESD53023.2021.9501725.

[11] Ou, X.—Li, H.: YNU_OXZ @ HaSpeeDe2 and AMI: XLM-RoBERTa with Ordered Neurons LSTM for Classification Task at EVALITA 2020. In: Basile, V., Croce, D., Maro, M., Passaro, L. C. (Eds.): EVALITA Evaluation of NLP and Speech Tools for

Italian. Proceedings of the Final Workshop (EVALITA 2020). Accademia University Press, Torino, Italy, 2020, pp. 102–109, doi: 10.4000/books.aaccademia.6912.

[12] Datta, A.—Si, S.—Chakraborty, U.—Naskar, S. K.: Spyder: Aggression Detection on Multilingual Tweets. Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, European Language Resources Association (ELRA), Marseille, France, 2020, pp. 87–92, https://aclanthology.org/2020.trac-1.14.

[13] Chakraborty, P.—Seddiqui, M. H.: Threat and Abusive Language Detection on Social Media in Bengali Language. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934609.

[14] Fersini, E.—Nozza, D.—Rosso, P.: AMI @ EVALITA2020: Automatic Misogyny Identification. In: Basile, V., Croce, D., Maro, M., Passaro, L. C. (Eds.): EVALITA Evaluation of NLP and Speech Tools for Italian. Proceedings of the Final Workshop (EVALITA 2020). Accademia University Press, Torino, Italy, 2020, pp. 21–28, doi: 10.4000/books.aaccademia.6764.

[15] Anzovino, M.—Fersini, E.—Rosso, P.: Automatic Identification and Classification of Misogynistic Language on Twitter. In: Silberztein, M., Atigui, F., Kornyshova, E., Métais, E., Meziane, F. (Eds.): Natural Language Processing and Information Systems (NLDB 2018). Springer, Cham, Lecture Notes in Computer Science, Vol. 10859, 2018, pp. 57–64, doi: 10.1007/978-3-319-91947-8_6.

[16] Bhattacharjee, A.—Hasan, T.—Ahmad, W.—Mubasshir, K. S.—Islam, M. S.—Iqbal, A.—Rahman, M. S.—Shahriyar, R.: BanglaBERT: Language Model Pretraining and Benchmarks for Low-Resource Language Understanding Evaluation in Bangla. In: Carpuat, M., de Marneffe, M. C., Meza Ruiz, I. V. (Eds.): Findings of the Association for Computational Linguistics: NAACL 2022. 2022, pp. 1318–1327, doi: 10.18653/v1/2022.findings-naacl.98.

[17] Sarker, S.: BanglaBERT: Bengali Mask Language Model for Bengali Language Understanding. 2020, https://github.com/sagorbrur/bangla-bert.

[18] Devlin, J.—Chang, M. W.—Lee, K.—Toutanova, K.: BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. CoRR, 2018, doi: 10.48550/arXiv.1810.04805.

[19] Reimers, N.—Gurevych, I.: Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. CoRR, 2019, doi: 10.48550/arXiv.1908.10084.

Sarif Sultan Saruar Jahan received his B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh. Currently, he is working as a Software Engineer at the BJIT Group. His research interests are NLP, image processing, deep learning, and transfer learning.

Raqeebir Rab received her B.Sc. degree in science with a major in computer science from the Augustana Faculty, University of Alberta, Canada in 2004. She completed her M.Sc. in computer science at the Concordia University, Montreal, Canada in 2012. She is an Assistant Professor at the Department of Computer Science and Engineering (CSE), Ahsanuallah University of Engineering and Technology (AUST), Dhaka, Bangladesh. Her research interests include wireless multihop networks (ad hoc and sensor networks) with an emphasis on mathematical modeling, performance analysis, protocol design, and data science.

Peom Dutta received his B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh. Currently, he is working as an Associate Data Analyst. His research interests include machine learning and artificial intelligence.

Hossain Muhammad Mahdi Hassan Khan received his B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh, in 2022. Currently, he is working as a Software Quality Assurance Intern. His research interests are machine learning, artificial intelligence, usage of deep learning in software testing, and cyber security.

Muhammad Shahariar Karim Badhon received his B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST), Dhaka, Bangladesh. Currently, he is working as a trainee at B-JET (Bangladesh-Japan ICT Engineers' Training Program). His research interests are machine learning and artificial intelligence.

Sumaiya Binte Hassan received her B.Sc. degree in cognitive systems in the cognition and brain stream from the University of British Columbia, Canada. Currently, she is working as a Research Assistant (RA) in the Brain, Attention, and Reality Lab, Canada. During her time as a Research Assistant, she focused mostly on VR and survey-based studies in the domain of evolutionary psychology.

Ashikur Rahman is a Professor in the Department of Computer Science and Engineering at Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh. He holds B.Sc. and M.Sc. degrees from BUET and his Ph.D. from the University of Alberta, Canada. He has worked as a post-doctoral researcher at various universities and his research focuses on cyber-physical systems, wireless networks, machine learning, and neural networks.