Sentiment Analysis of Facebook and YouTube Bengali Comments Using LSTM and Bi-LSTM

Nolak Kapali
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh nolak15-12029@diu.edu.bd

Tariquzzaman Tuhin
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh tariquzzaman15-11912@diu.edu.bd

Anik Pramanik
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
anik15-9495@diu.edu.bd

Md.Sadekur Rahman
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
sadekur.cse@daffodilvarsity.edu.bd

Dr. Sheak Rashed Haider Noori
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
drnoori@daffodilvarsity.edu.bd

***Abstract*—Recently, sentiment analysis has been performed on various information on social media to derive market intelligence. As we know social media is filled with different content as a result audiences are interacting there making it a huge opportunity to perform sentiment analysis on this information. In terms of Bengali content, many audiences on social media interact with the Bengali language which makes social media a treasure trove to perform sentiment analysis in Bengali Natural Language Processing (NLP) field. Here in this study sentiment analysis has been performed on the audience's Bengali comments expressing different views towards social media's Bengali contents. The dataset contains 4000 Bengali comments collected from Facebook and YouTube Bengali Contents. Here positive, negative, and neutral classes are used to categorize the Bengali data, and a tokenizer from the Keras library is used to tokenize the Bengali text. Deep learning algorithm Long Short-Term Memory (LSTM) and BiDirectional Long Short-Term Memory (Bi-LSTM) are performed and Bi-LSTM has the highest accuracy of 97.25% than LSTM.**

Keywords—Sentiment Analysis, Bengali Contents, Social Media, NLP, LSTM, Bi-LSTM.

I. Introduction

Social media is blasting with information these days. Contents can be seen in social media. A content is any kind of videos or texts or photos etc. that a person or group of people posts to interact with the audiences. These contents can be videos or photos of product advertisement or blogs of an influencer or simple plain texts of any product reviews which means it can be anything. These contents in social media is increasing each day making some social media such as Facebook and YouTube a trending place to interact with audiences. As Facebook and YouTube popularity is increasing every day, Bengali social media content is also increasing in these platform filled with Bengali audience reaction towards the Bengali contents. In Bengali contents we can see from influencers, content creators to business and entertainment brands producing contents to promote their brands to increase their profit and to gain more followers and these all are Bengali contents of social media filled with huge public reactions expressed as Bengali comments. Social media is full of data and information, sentiment analysis can be performed on this vast information to know the opinions behind each and every people. As we know sentiment analysis is basically the analyzation of all the sentiments of audiences towards particular topics. In the world filled with so many different languages Bengali language is falling behind in research work of sentiment analysis due to the complexity and poor resources [1]. That's why in this study, sentiment analysis has performed on Bengali comments which are the opinions of audience's reaction toward Bengali contents of social media. The dataset is filled with 4000 Bengali comments collected from Bengali contents of Facebook and YouTube and we are performing sentiment analysis using these social media content's comments. Deep learning algorithm Long Short Term Memory (LSTM) and Bi Directional Long Short Term Memory (Bi-LSTM) are performed here. The reason for doing this study are follows:

- Generate the insider views of public towards the Bengali contents.
- Create a publicly available Bengali dataset. The scarcity of Bengali dataset is high as we face this during our study.
- Build a better deep learning model for sentiment analysis on Bengali language.

Our contributions are as follows:

- A deep learning model that classifies sentiments from Bengali text.
- A publicly available dataset of 4000 Bengali comments with labelled 3 class sentiments.

II. Related Works

Sentiment analysis has performed on many languages but our study focuses on Bengali language. Below some researchers done sentiment analysis work in Bengali language are studied and discussed here.

This paper shows bilingual sentiment classification where LR, ET, RF, SVM, RR, LSTM classifiers are performed on Bengali and its machine translated English corpora to compare the performance and positive, negative and neutral class was balanced using SMOTE which shows better performance increasing the F1 scores of the classifiers and also it shows that machine translated English corpus has better classifier performance [1].

To detect sentiment polarity of Bengali tweets this paper use 1500 Bengali tweets where SentiWordNet lexical knowledge is used to augmented tweet word with sentimental tag found in the dictionary. Proposed model is built with CNN and compare it with deep belief network (DBN) where CNN outperforms DBN with 46.80% accuracy [2]. To detect emotions from Bengali text this paper proposed an automated system which detects happy, sad, tender, scared, excited, angry emotions from 7500 Bengali textual data. The proposed model uses NB and Topical approach on both sentence and document level where topical approach outperforms NB with 90% on sentence level [3].

This paper present a depression dataset where the dataset filled with social media Bengali blogs which are categorized into 2 emotion classes happy and sad. Sentiment classification run in document level where it classifies if a post is happy or sad using ML algorithms MNB, RF, DT, linear SVC where MNB get the best accuracy of 86.67% [4]. This paper shows sentiment analysis on manually created 1601 Bengali comments dataset with praise, criticism and sadness emotions and another dataset with 2979 Bengali comments with positive, negative and neutral sentiments.

SVM, DT and MNB are performed where SVM outperformed them all by 64.596% accuracy in the emotion dataset and 73.490% in sentiment dataset. [5].

This paper use 7905 Bengali product reviews dataset and where positive, negative and neutral class sentiments added manually to the dataset according to the user ratings of product review and oversampling was done to balance the classes. SVM, KNN, LR, RF, XGB algorithm all with 10 cross-validation has run on the model where KNN has highest 96.25% accuracy [6].

BanglaSenti is a lexicon-based corpus or dataset of 61582 Bengali words including positive, negative and neutral labelling. By collecting data from SentiWordNet they transform the English words consisting positive, negative and neutral words into Bengali words with positive, negative and neutral words [7].

This paper with 2000 Bengali book review corpus detects sentiment polarity using tf-idf values with n-gram features. LR, MNB, RF, DT, KNN, SVM, SGD all with 10-fold crossvalidation performed but MNB has highest 84% accuracy [8]. This paper shows sentiment analysis of online restaurant review with 1000 Bengali review. Ml algorithms DT, RF, MNB all with K fold cross-validation performed and MNB gave highest 80.48% accuracy with 6-fold cross validation
[9].

This paper creates a benchmark dataset with 12000 Bengali reviews from YouTube. They translated the Bengali data to English data and supervised ml algorithms LR, SVM, RF, ET and unsupervised lexicon-based approach using VADER, TextBlob and SentiStrength used both on Bengali and English dataset to classify the sentiments. Used cross-domain dataset for transfer learning and SVM performs best with 93% accuracy in Bengali and 93.5% accuracy in English corpus [10].

This paper shows sentiment polarity detection in Bengali tweets where the polarities are positive, negative and neutral. Dataset of 1500 Bengali tweets, SentiWordNet is used to augment the twitter dataset words with sentimental tag found in SentiWordNet. As the dataset is small for better performance 10-fold cross validation is performed with LSTM and has the best accuracy 55.27% [11].

This paper presents sentiment polarity detection of Facebook 10819 Bengali data. SVM, NB, DT, AdaBoost, RF, LSTM and CNN used here to detect positive and negative sentiments where classical and deep learning algorithm are compared. LSTM has highest 96.95% [12].

This paper proposed a Bi-LSTM model that detects sentiment polarity and categorized them into positive and negative from Bengali Facebook data. SVM, DT, Logistic Linear Regression and also RNN with Bi-LSTM which is the proposed model has performed and RNN with Bi-LSTM outperformed them all by 85.67% accuracy [13]. This paper shows sentiment polarity detection in Bengali 7000 Facebook dataset using doc2vec model with 2 sentiment class positive and negative. SGD, DT, KN classifier, LDA, GausssianNB, SM, Bi-LSTM has performed on these dataset and Bi-LSTM outperformed them all by 77.85% accuracy [14].

This paper introduces a Bengali dataset named BAN-ABSA with 9009 comments manually annotated with sport, politics, religions and others aspects with their own polarity of positive, negative and neutral class. SVM, CNN, LSTM, BiLSTM are performed where Bi-LSTM has highest 78.75% accuracy in aspect extraction but in sentiment classification it has 71.08% accuracy [15].

This paper shows sentiment polarity detection on Bengali tweets data using character n gram features with MNB algorithm. They have used SentiWordNet and their suggested model with MNB has 48.5% accuracy [16].

This paper presents a modified Bengali VADER which can classify sentiment from Bengali language without translating the data into English and they developed Bengali polarity lexicon from VADER English lexicon. Also stemming, Bengali boosting words are added and also they bi-gram and tri-gram are used to improve the performance. [17]. This paper presents multilingual BERT with the approach of transfer learning in Bengali sentiment analysis. Two Bengali datasets where the first one is two class positive and negative sentiment and the other on is three class positive, negative, neutral sentiment. Three model one with BERT, another with word2vec and another one with fastText embedding proposed but BERT performs best with 71% and 60% accuracy for 2 and 3 class sentiments respectively [18].

This paper proposes a new approach using word2vec for sentiment analysis of Bengali 16000 comments. They performed word embedding approach and word2vec is one of the techniques and their proposed approach is with word2vec combining the two information where the first one is the dataset contextual information and second one is the sentiment information of the dataset. To create the new approach, they have built a list containing some highly positive and negative words put with the polarity score of each word is created and the accuracy is achieved to 75.5% [19].

In our work we wanted to build a better deep learning model and if we follow these we can see that our model accuracy is better than some of the past work’s [11] [13] [14] [15] deep learning model accuracy of their sentiment analysis task. In the next methodology section we discussed our sentiment analysis work on Bengali comments in details.

III. METHODOLOGY

This section describe all the procedure to build the models based on LSTM and Bi-LSTM algorithm of our sentiment analysis task. Fig. 1 shows the working procedure of methodology and all the steps to execute it are discussed below. Python is used as the programming language and it was implemented using Google Colaboratory.

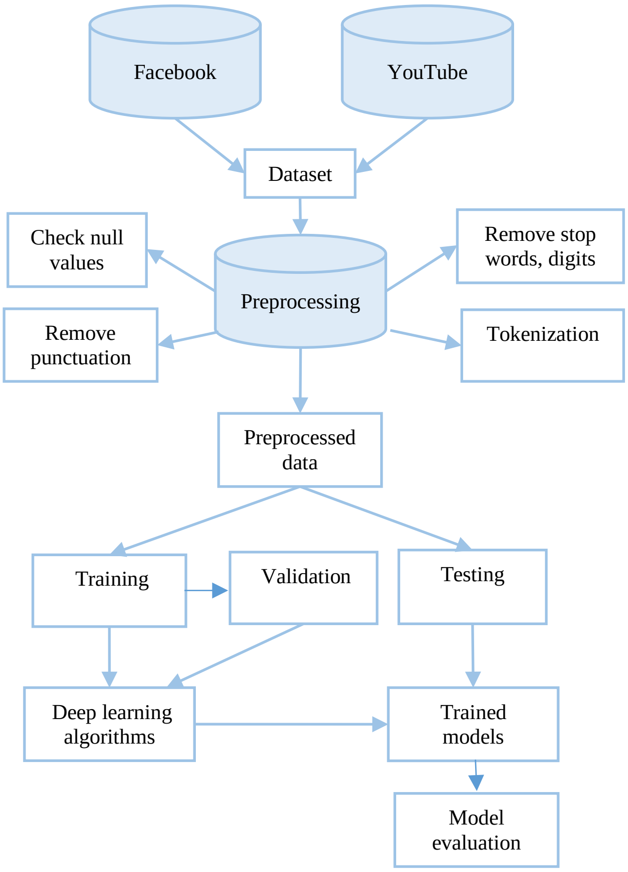

Fig. 1. Flow-chart of the working procedure

A. Data Collection

We have collected 4000 audience's Bengali comments from Facebook and YouTube Bengali contents. We have used google sheet to manually collect our data and manually categorized our data into three sentiments named positive, negative and neutral. Table. I and Fig. 2 shows the distribution of comments between the three classes.

TABLE I. DISTRIBUTION OF COMMENTS INTO THREE CLASSES

Total Comments	Positive	Neutral	Negative
4000	1244	1238	1518

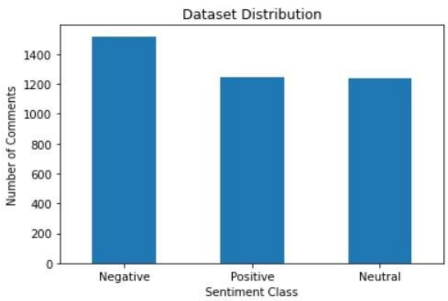

Fig. 2. Dataset distribution

B. Data Preprocessing

For building the machine learning model successfully and fitting Bengali data into that model data preprocessing is a must [4]. Data preprocessing has two parts here which are data cleaning and tokenization.

1) Data Cleaning

In data cleaning, first we removed the punctuation from the google sheets, removed stop words using BNLP corpus and digits using regular expression and before all of that we have checked if there are any null values available in the dataset and deleted it accordingly.

2) Tokenization

Tokenization is the process where text is broken into chunks of words or characters or sentence etc. and all of these chunks are called tokens. For tokenization, we have used Keras preprocessing module's tokenizer class which tokenizes our Bengali text and convert each token or word into integers and whole sentence into sequence of integers values. As we are using deep learning algorithm LSTM and Bi-LSTM so both needed sequence data to perform and that's why Keras tokenizer class is used here. In the tokenizer class fit_on_text and text_to sequence method was used to convert the tokens into sequence of integers. The first method updates the vocabulary according to the text it gets from the dataset and gives vocabulary's each word a word index or an integer number based on the frequency of that word. Suppose dataset is "I am a girl. I love ice-cream. I am awesome" so according to this method it gives integer values basis on the word frequency. So, after applying this method it will be I =1, am=2, a=3, girl=4, love=5, ice-cream=6, awesome=7. So here I occur more than am that's why it gets more priority here. So, in this dictionary every unique word has given a number. In the second method if we follow the first example then after applying first method all the words index are I =1, am=2, a=3, girl=4, love=5, ice-cream=6, awesome=7 and if second method is applied then the result is [[1,2,3,4], [1,5,6], [1,2,7]] which represents the sequence of integers and the integers seat together according to the original sentence word sequence. So the second methods puts the tokens into sequence of integers and so from generating tokens to sequence of integers from it and all the steps are shown in fig. 3.

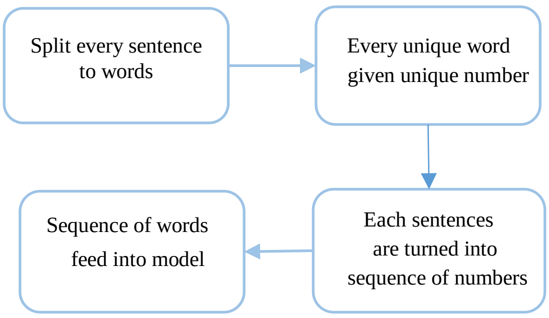

Fig. 3. Flow-chart of tokenization process

After converting every text into a sequence of numbers it is padded to fit into same lengths and later feed to both LSTM and Bi-LSTM model.

C. Model Development

For classifier, we have selected neural networks rather than classical one because neural network performs better than classical or traditional classifier. Two deep learning algorithms LSTM and Bi-LSTM are used to perform sentiment analysis task where first model is based on LSTM network and second model is based on Bi-LSTM and Keras layers are used to build these models in Google Colaboratory.

LSTM: LSTM is an advance form of recurrent neural network (RNN). It remembers a data longer than RNN and also using it's 3 gates it decides which data is more important and which data it needs to forget and that's why for all the sequence data related task LSTM is suited better than RNN. For this study LSTM uses some layers to execute the classification task and is described below.

1. *Embedding Layer:* It is the first hidden layer of this neural network. Here all the integer encoded input dimension will be reduced to make the model more efficient.
2. *LSTM Layer:* This next layer is LSTM layer which has 16 nodes with a dropout of 0.2.
3. *Dense Layer/Output Layer:* It is a fully connected neural network layer where it's 3 nodes are connected to the pervious neurons of the previous layers. It is an output layer with softmax activation function.

Bi-LSTM: Bi-LSTM is Bidirectional LSTM network which is an advance version of LSTM. Bi-LSTM is built with two LSTM network where one network will take the input in forward way and the other in backward way and so information flows from backward to forward and forward to backward and that's why it understands the context better [13]. All the layers that are used to build the Bi-LSTM sequential model are discussed below.

1. *Embedding Layer:* It is first layer in this where input is integer coded input and this layer takes the input and reduced the dimension of it and pass it to the next layer.
2. *Bidirectional LSTM Layer:* This hidden layer has 16 nodes and it has a dropout rate of 0.2
3. *LSTM Layer:* This layer is a LSTM layer with 16 nodes with a dropout of 0.2.
4. *Dense Layer/Output Layer:* This layer has 3 nodes and these nodes are fully connected with the previous layers. This is an output layer and as there are three sentiments the number of nodes here are 3. In this layer for activation function, softmax is used.

IV. Experimental Result and Analysis

For experimental setup, the dataset is divided into 80% training set and 20% testing set. The training set was split into 10% validation set with a validation split for both LSTM and Bi-LSTM model. Table II shows the proper distribution of the training, validation and testing set for both LSTM and Bi-LSTM model.

TABLE II. DISTRIBUTION OF TRAINING, VALIDATION AND TESTING SETS

Models	Training (%)	Validation (%)	Testing (%)
For both LSTM and Bi-LSTM model	70%	10%	20%

A. Hyper-Parameters Tuning

For training the data, the hyper parameter of the model need to be tuned to generate a good result. For both LSTM and Bi-LSTM model, in the embedding layers input dimension is set to 5000, output dimension is set to 64, in the output layers softmax activation function is used and for optimizer adam is used and for loss function as it is multiclass classification categorical crossentropy is used. In the output layers nodes are kept to 3 for both models as our sentiment analysis is based on three class sentiments. Data is trained on 40 epoch with batch size of 12 for both models. For LSTM model, a single LSTM layer is used with 16 nodes and for Bi-LSTM model a single Bi-LSTM and LSTM layer is used with 16 nodes on both layers.

B. Evaluation Result

After training the data and tuning the hyper-parameters to get more precise results on both models, experimental results are discussed and analyzed in this section. Table III shows the comparison of both model's accuracy.

TABLE III. COMPARISON OF MODEL'S ACCURACY

Model	Accuracy (%)
Model Based on LSTM	96.88%
Model Based on Bi-LSTM	97.25%

In Table III Bi-LSTM shows better accuracy than LSTM. To analyze the results more, graphs of accuracy and loss for both models and classification tables are discussed below. Classification tables are values of precision, recall, f1 score performs on different classes of dataset which helps us analyze the algorithm better [9].

Fig. 4 and Fig. 5 shows both LSTM and Bi-LSTM model accuracy and loss in the time of training and testing the dataset after every epoch.

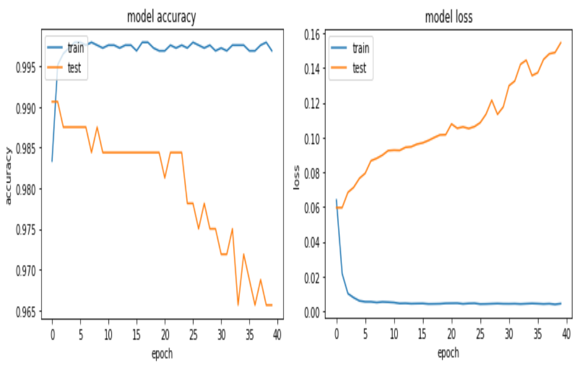

Fig. 4. LSTM model accuracy and loss

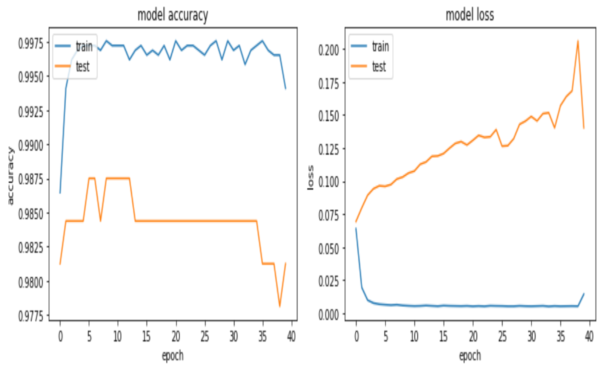

Fig. 5. Bi-LSTM model accuracy and loss

Table IV and Table V shows the classification table for LSTM and Bi-LSTM where the values of precision, recall, f1score and support values are shown on the three sentiment class.

TABLE IV. CLASSIFICATION TABLE FOR LSTM

Sentiment	**Precision**	**Recall**	**F1 Score**	**Support**
Negative	0.97	0.97	0.97	318
Neutral	0.98	0.95	0.97	242
Positive	0.96	0.97	0.97	240

TABLE V. CLASSIFICATION TABLE FOR BI-LSTM

Sentiment	**Precision**	**Recall**	**F1 Score**	**Support**
Negative	0.98	0.98	0.98	296
Neutral	0.98	0.96	0.97	270
Positive	0.95	0.97	0.96	234

According to the classification report, LSTM model negative class f1 score is 0.97, positive is 0.97 and neutral class is 0.97 which means the model can correctly tell that 97% data are negative, 97% are positive and 97% are neutral. For Bi-LSTM model negative, neutral and positive f1 score are 0.98, 0.97 and 0.96 respectively which means that Bi-LSTM model can correctly classify negative data are 98%, neutral data are 97% data and positive data are 96%.

In the below table VI shows the confusion matrix of LSTM algorithm and table VII shows the confusion matrix of Bi-LSTM algorithm. Confusion matrix shows how accurately the model predicts among all the classes. Here in this LSTM and Bi-LSTM confusion matrix actual and predicted class are given for each positive, negative and neutral sentiments.

TABLE VI. CONFUSION MATRIX OF LSTM

		Predicted class		
	Sentiment	***Negative***	***Neutral***	***Positive***
Actual class	***Negative***	310	2	6
	Neutral	7	231	4
	Positive	4	2	234

TABLE VII. CONFUSION MATRIX OF BI-LSTM

		Predicted class		
	Sentiment	***Negative***	***Neutral***	***Positive***
Actual class	***Negative***	290	2	6
	Neutral	3	260	7
	Positive	4	2	228

From the confusion matrix of both LSTM and Bi-LSTM it shows that in 800 test data, Bi-LSTM predicts correctly more than LSTM.

V. CONCLUSION AND FUITURE WORK

In this study, sentiment analysis has been performed on Bengali comments collected from social media Bengali contents. These comments are the reaction of audiences towards different Bengali contents. The dataset is consists of these 4000 Bengali comments taken from Facebook and YouTube posts. Here in this study, we have performed deep learning algorithm LSTM and Bi-LSTM for the sentiment analysis task where Bi-LSTM give better performance than LSTM with 97.25% accuracy. This study focuses on Bengali language as there are few works have been done in this field. Using our work, Bengali content creator can know their audience opinions about their contents without reading all the comments and researcher can use this dataset for their research work.

As the dataset is manually created, finding Bengali data is challenging in social media and that is why we can see that the dataset that we build is not so large which is one of the limitations and so for future improvements we will increase the size of the dataset. One of the limitation that we face the dataset we have collected and in there some comments are written in English words but in Bengali letters also some has missing letters or incorrect spelling of Bengali words and these word does not produce any proper meaning in Bengali language and that's why in future we want to improve this issue. Also we want to increase our horizon in social media platform that's why we will add Bengali data from TikTok and Instagram Bengali contents. As the Bengali comments are collected from Bengali contents we find some comments that are in Bengali language but written in English words also some post that are heavily one sentiment based can create class misbalanced which can cause class misbalancing issues to our model and we want to work on these issues in our future improvements. We will also use pre-trained word embedding to both model in the future work.

ACKNOWLEDGMENT

We are immensely grateful to the Virtual Multidisciplinary Research Lab (VMRL) colleagues who shared their expertise and knowledge with us throughout this research and consider assisting us with their insight and experience.

REFERENCES

[1] S. Sazzed and S. Jayarathna, “A sentiment classification in bengali and machine translated english corpus”, In 2019 IEEE 20th international conference on information reuse and integration for data science (IRI), IEEE, pp. 107-114, July 2019.

[2] K. Sarkar, “Sentiment polarity detection in Bengali tweets using deep convolutional neural networks”, Journal of Intelligent Systems, 28(3), pp.377-386, 2019.

[3] 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 C.omputer and Communication Systems (ICCCS), IEEE, pp. 360-364, February 2019.

[4] M.R.H. Khan, U.S. Afroz, A.K.M Masum, S. Abujar and S.A. Hossain, “Sentiment analysis from bengali depression dataset using machine learning”, In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-5, July 2020.

[5] S.A. Mahtab, N. Islam and M.M. Rahaman, “Sentiment analysis on bangladesh cricket with support vector machine”, In 2018 international conference on Bangla speech and language processing (ICBSLP), IEEE, pp. 1-4, September 2018.

[6] M.T Akter, M. Begum and R. Mustafa, “Bengali Sentiment Analysis of E-commerce Product Reviews using KNearest Neighbors”, In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), IEEE, pp. 4044, February 2021.

[7] H. Ali, M.F. Hossain, S.B. Shuvo and A. Al Marouf, “Banglasenti: A dataset of bangla words for sentiment analysis”, In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-4, July 2020. [8] E. Hossain,O. Sharif, and M. Moshiul Hoque, “Sentiment polarity detection on Bengali book reviews using multinomial naive bayes”, In Progress in Advanced Computing and Intelligent Engineering , Springer, Singapore, pp. 281-292, 2021.

[9] O. Sharif, M.M Hoque and E. Hossain, “Sentiment analysis of Bengali texts on online restaurant reviews using multinomial Naïve Bayes”, In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, pp. 1-6, May 2019.

[10] S. Sazzed, “Cross-lingual sentiment classification in low-resource bengali language”, In Proceedings of the sixth workshop on noisy usergenerated text (W-NUT 2020), pp. 50-60, November 2020.

[11] K. Sarkar, “Sentiment polarity detection in Bengali tweets using LSTM recurrent neural networks”, In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), IEEE,pp. 1-6, February 2019

[12] P. Chakraborty, F. Nawar and H.A. Chowdhury, “Sentiment Analysis of Bengali Facebook Data Using Classical and Deep Learning Approaches”, In Innovation in Electrical Power Engineering, Communication, and Computing Technology, Springer, Singapore, pp. 209-218, 2022.

[13] A.A. Sharfuddin, M.N. Tihami and M.S.Islam, “A deep recurrent neural network with bilstm model for sentiment classification”, In 2018 International conference on Bangla speech and language processing (ICBSLP), IEEE, pp. 1-4, September 2018.

[14] M.T. Hoque, A. Islam, E. Ahmed, K.A Mamun and M.N. Huda, “Analyzing performance of different machine learning approaches with doc2vec for classifying sentiment of bengali natural language”, In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp. 1-5, February 2019.

[15] M. Ahmed Masum, S. Junayed Ahmed, A. Tasnim and S. Islam, “BAN-ABSA: An Aspect-Based Sentiment Analysis Dataset for Bengali and Its Baseline Evaluation”, In Proceedings of International Joint Conference on Advances in Computational Intelligence, Springer, Singapore, pp. 385-395, 2021.

[16] K. Sarkar, “Using character n-gram features and multinomial naïve bayes for sentiment polarity detection in Bengali tweets”, In 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), IEEE, pp. 1-4, January 2018.

[17] A. Amin, I. Hossain, A. Akther, and K.M. Alam, “Bengali vader: A sentiment analysis approach using modified vader”, In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp. 1-6, February 2019.

[18] K.I. Islam, M.S. Islam and M.R. Amin, “Sentiment analysis in Bengali via transfer learning using multi-lingual BERT”, In 2020 23rd International Conference on Computer and Information Technology (ICCIT), IEEE, pp. 1-5, December 2020.

[19] M. Al-Amin, M.S. Islam and S.D. Uzzal, “Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words”, In 2017 international conference on electrical, computer and communication engineering (ECCE), IEEE, pp. 186-190, February 2017.