Sentiment analysis on Bangla conversation using machine learning approach

Mahmudul Hassan[1], Shahriar Shakil[1], Nazmun Nessa Moon[1], Mohammad Monirul Islam[1], Refath Ara Hossain[1], Asma Mariam[1], Fernaz Narin Nur[2]

[1]Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Faculty of Science and Information Technology, Notre Dame University, Dhaka, Bangladesh

Article Info

ABSTRACT

Nowadays, online communication is more convenient and popular than face-to-face conversation. Therefore, people prefer online communication over face-to-face meetings. Enormous people use online chatting systems to speak with their loved ones at any given time throughout the world. People create massive quantities of conversation every second because of their online engagement. People's feelings during the conversation period can be gleaned as useful information from these conversations. Text analysis and conclusion of any material as summarization can be done using sentiment analysis by natural language processing. The use of communication for customer service portals in various e-commerce platforms and crime investigations based on digital evidence is increasing the need for sentiment analysis of a conversation. Other languages, such as English, have well-developed libraries and resources for natural language processing, yet there are few studies conducted on Bangla. It is more challenging to extract sentiments from Bangla conversational data due to the language's grammatical complexity. As a result, it opens vast study opportunities. So, support vector machine, multinomial naïve Bayes, k-nearest neighbors, logistic regression, decision tree, and random forest was used. From the dataset, extracted information was labeled as positive and negative.

Corresponding Author:

Nazmun Nessa Moon
Department of Computer Science and Engineering, Daffodil International University
Dhaka 1207, Bangladesh
Email: moon@daffodilvarsity.edu.bd

1. INTRODUCTION

People have conversations in their daily life. People express their feelings and opinions in their conversations. These feelings and opinions can be categorized into sad, anger, happy, worried, disgusted, frightened, complement, motivation, suggestions, and neutral [1]. To detect subjective information such as opinions, attitudes, and feelings expressed in text Sentiment analysis or opinion mining aims to use automated tools [2]. In our research work we merged them into two main categories of positive and negative [3]. Sentiment analysis can be done by capturing both semantic and sentiment similarities among words [4]. Our model can identify whether a part of any conversation is positive or negative. These two categories expose the sentiment of the people who said it. Analyzing sentiment from people's speech is a tough job because in a single sentence people can express various types of sentiment at the same time. Only the people who listen to it, can understand the sentiment properly. Our proposed model can extract sentiment from

people's conversation with a closer accuracy of real life. In this research work we proposed a model that can extract sentiment from conversation as positive or negative sentiment. To pursue that we split our dataset into 80:20 ratio. For training purposes, we used 80% data and for testing purposes we used 20% data. It helps to increase the accuracy of the model. Based on the training dataset the accuracy of the model fully depends on the training dataset. We have used some techniques such as changing the parameters of machine learning models to get more accurate results. We achieved about 86% accuracy on the support vector machine. Rest of the algorithms perform closely to the highest accuracy.

2.   LITERATURE REVIEW

Extracting sentiment from Bangla conversational data is a method for determining if a conversation is positive or negative. Bhowmik *et al.* [5] developed deep learning models for Sentiment analysis on Bangla text using an extended lexical data set. They employed the rule-based Bangla text sentiment score system to extract polarity from large texts. These polarities, along with the pre-processed text, are then used as training samples by the neural network. The pre-processed texts are displayed as a vectorization of words derived from pre-trained word embedding models with various word counts. A Word2Vec matrix containing the top highest probability word is used as a weighted matrix on the embedding layer to fit the deep learning models. This paper also includes a thorough examination of selective deep learning models, as well as some fine-tuning. Their proposed hierarchical approach was accurate to the tune of 78.52 percent, 80.82 percent, and 84.18 percent, respectively. According to Aurpa *et al.* [6] certain items, such as threats and sexual harassment, were more accessible than traditional media. Harassment, vulgarity, personal assaults, and bullying can all occur because of extremely toxic internet content. Bangla's use of Facebook has risen in recent years due to its status as the world's seventh most spoken language. The use of offensive comments in Bangla on Facebook has also grown significantly, but there is little research on the subject. They focus on recognizing abusive Bangla language remarks on social media (Facebook) that can be filtered out in the early phases of social media attachment in this study. To classify hostile comments quickly and accurately, transformer-based deep neural network models were used. They employed pre-training language architectures bidirectional encoder representations from transformers (BERT) and efficiency learning an encoder that accurately classifies token replacements (ELECTRA). The average accuracy, precision, recall, and f1-score were used to assess the proposed models. The results have revealed that our BERT and ELECTRA architectures are performing admirably, with test accuracy of 85.00 percent and 84.92 percent, respectively. Rahib *et al.* [7] conducted this study to investigate how Bangladeshis are reacting to and dealing with the coronavirus disease (COVID-19) scenario. In this investigation, the status and comments on COVID-19 concerns were gathered from multiple Facebook pages and YouTube channels run by reputable Bangladeshi news organizations and health specialists. Throughout the study, a variety of machine learning algorithms were studied, ranging from conventional algorithms like support vector machine and random forest to deep learning algorithms like convolutional neural networks and long short-term memory. Experiments were carried out on a 10,581-data-point categorized data set belonging to the authors. When evaluating the performance of various models in terms of model assessment, the results demonstrate that long short-term memory exceeds all of them, with an accuracy of 84.92 percent. To detect the polarity of textual Facebook posts in Bangla containing people's points of view on Bangladesh Cricket, Faruque *et al.* [8] proposed a sentiment polarity detection approach that uses three popular supervised machine learning algorithms: naive Bayes (NB), support vector machines (SVM), and logistic regression (LR). With an accuracy of 83 percent when considering n-gram as a feature, LR outperformed SVM and NB. Iqbal *et al.* [9] proposed a four-step process for categorizing six emotions in Bengali literature, including data crawling, pre-processing, labelling, and verification, with 7,000 texts labeled into six basic emotion groups. The dataset is graded with a score of 0.969. Cohen's score reflects the close collaboration between corpus annotators and experts. According to the analysis of appraisal, the distribution of emotion words also follows Zipf's law. The BEmoC study's findings were also presented in terms of coding consistency, emotion density, and the most utilized emotion words.

Shetu *et al.* [10] established a paradigm for parsing text data in paragraphs. To extract sentiment from a text, they employed the bag of words method and lexical analysis method. Mamun *et al.* [11] demonstrated that the ensemble approach (i.e., *logistic regression+random forest+support vector machine*) with frequency-inverse document frequency (*unigram+bi-gram+trigram*) features outperformed the other classifier models on the developed dataset, achieving the highest accuracy of 82 percent. Most of the emotions conveyed on social media platforms are expressed through writing (such as status, tweets, comments, and reviews). presents an ensemble-based method for categorizing Bengali textual sentiment into positive and negative categories. Because the Bengali sentiment corpus was unavailable, this effort additionally created a dataset called "Bengali sentiment analysis dataset". Neethu and Rajasree [12] attempted to assess the sentiment of Twitter posts in a particular domain. They suggested a new feature

vector that can differentiate between positive and negative sentiment in tweets. In order to examine twitter data for sentiment analysis, Jain and Dandannavar [13] used naive Bayes and decision tree machine learning methods. Because it is scalable and fast, their proposed model employs Apache Spark. Rahman and Dey [14] provide two freely accessible Bangla datasets for sentiment analysis based on aspects. One dataset contains user comments regarding cricket that have been human-annotated, while the other features restaurant customer reviews. They also presented a fundamental method for analyzing our datasets utilizing the aspect category extraction subtask.

3.   RESEARCH METHOD

Research section will illustrate the overall architecture of our proposed system. The research method is listed in Figure 1 as data collection, data pre-processing, model selection, statistical analysis, and its implementation will be discussed in this portion. In Figure 1 the full method at a glance is shown.
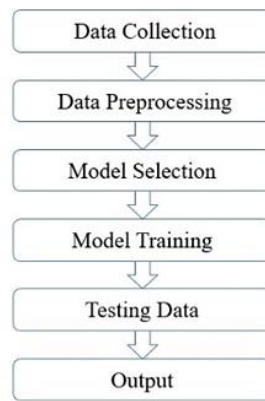

Figure 1. Method at a glance

3.1. Data collection procedure

From various Bangla movies and short film scripts, we collected conversation data for our research work. These conversations covered a large scale of topics like food, family, motivation, fraud, business, and friends. After analyzing those collected data, we will split it into two categories: positive and negative. We have collected about 1,141 data. These conversations include emotions like happy, sad, anger, worried, and afraid. These categories help us to differentiate the whole dataset into two main categories of Positive and Negative. Among 1,141 data there was 570 data for positive sentiment and for negative it was 571 data. Figures 2 and 3 shows the sample dataset.
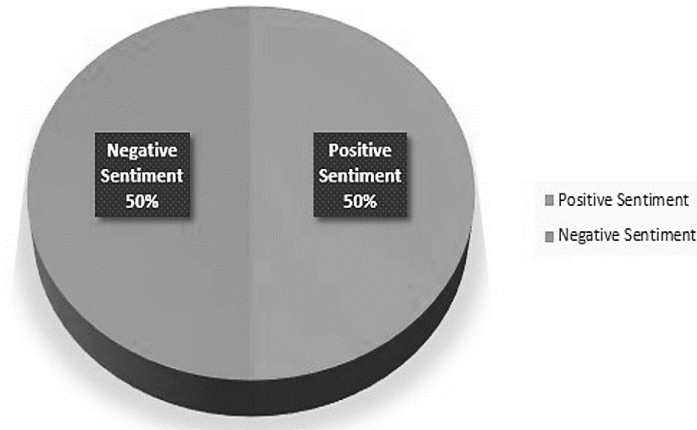

Figure 2. Sample data

Figure 3. Class label distribution

3.2. Data preprocessing and organizing

Firstly, we collect data from scripts and store them into an xlsx file. The dataset we have collected has two attributes. These are positive and negative. As we already discussed, we collect data from movie and short film scripts as conversation. Every conversation starts with a single word or single sentence. People can express their feelings, emotions, and thoughts through a single word or sentence. To classify these expressions into two main attributes we merged happiness, joy, motivation, and thankfulness into positive conversations, and for negative conversation we merged sad, anger, backbiting, and worries. During pre-processing, we remove punctuation in the first step. In natural language processing, for every language, it is essential to identify and remove stop words. For our research work, we have collected Bangla stop words and removed them to clean our data. There were about 410 stop words in the Bangla language. For example: 'অতএব', 'অথচ', 'এই', 'একই', 'একটি', 'হয়', 'হয়তো', 'কিন্তু', 'কী', and 'কে'. Here, Figure 4. shows the python code for removing Bangla stop words and punctuations and Figure 5. shows the cleaned data what we pre-processed.

```
[7]: def process_conversations(Conversation):
         stp = open('bangla_stopwords.txt','r',encoding="utf8").read().split()
         result = Conversation.split()
         Conversation = [word.strip() for word in result if word not in stp ]
         Conversation =" ".join(Conversation)
         Conversation = re.sub('[^\u0980-\u09FF]',' ',str(Conversation))
         return Conversation
```

Figure 4. Removing stop words and punctuations

```
Original:
দুর্দান্ত তুমি কি তাতে খুশি?  হ্যাঁ, আমি শিহরিত খুঁজে পেয়েছি!
Cleaned:
দুর্দান্ত খুশি  হ্যাঁ  শিহরিত খুঁজে পেয়েছি
Sentiment:-- positive

Original:
আমরা স্বামী-স্ত্রী এখনো একে অপরকে ভালবাসি
Cleaned:
স্বামী স্ত্রী এখনো অপরকে ভালবাসি
Sentiment:-- positive

Original:
আপনি যা করেছেন তার জন্যে আপনাকে মেরে ফেলা উচিত
Cleaned:
জন্যে মেরে ফেলা
Sentiment:-- negative
```

Figure 5. Cleaned data

To extract features from each of the conversations, several words and a number of characters are needed. Figure 6 shows the result, respectively. After preprocessing procedure label encoding method applied to the sentiment column. And then a pickle file generated. The pickle file contains temporary data for reuse and also saves time during runtime execution. In this work, our cleaned data is stored as a pickle file for upcoming procedures. We need to demonstrate our dataset data where highlights are age, occupation, house type, want to switch jobs and we are giving low highlighting to other attributes. In Figure 7, cleaned data along with counts of each conversation length and character is shown.
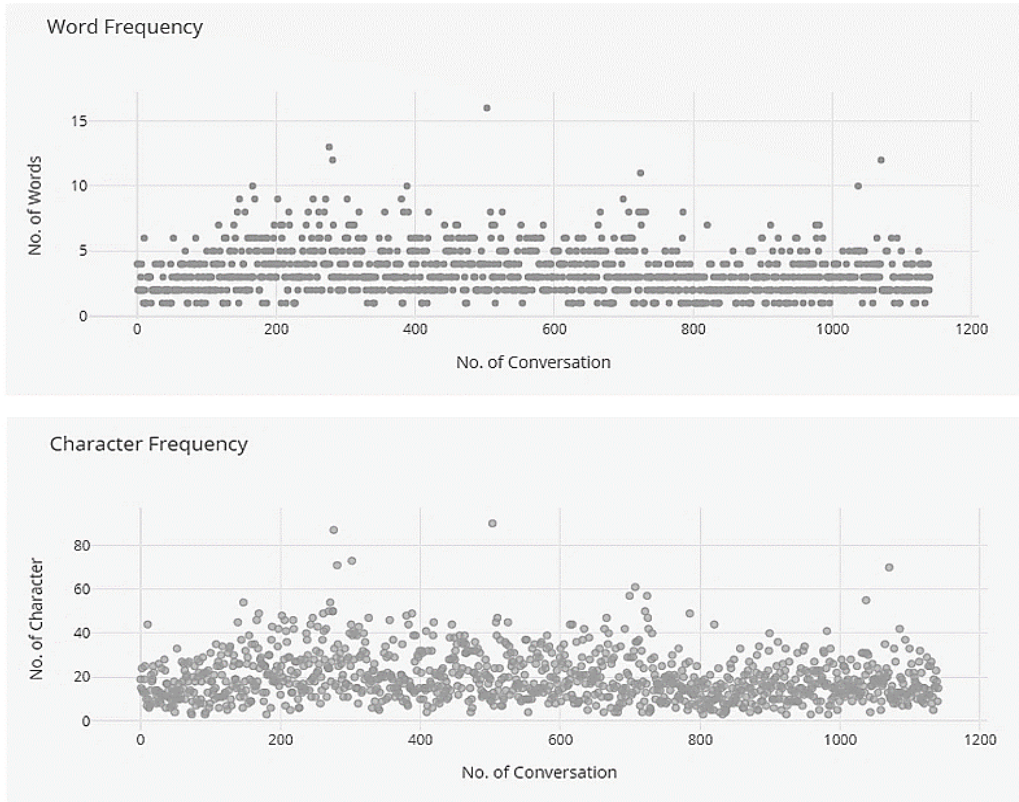

Figure 6. Word frequency and character frequency

	Conversation	Sentiment	cleaned	length	no_char
0	আমি বুথ থেকে কল দেয়ার জন্য দুঃখিত	negative	বুথ কল দেয়ার দুঃখিত	4	19
1	আমি তোমার সম্পর্কে দুঃখিত	negative	সম্পর্কে দুঃখিত	2	15
2	আমার ভয় হচ্ছিল আমি খুব ভয় পেয়েছিলাম	negative	ভয় হচ্ছিল ভয় পেয়েছিলাম	4	24
3	অবশ্যই ধন্যবাদ	positive	অবশ্যই ধন্যবাদ	2	14
4	কী সুন্দর!! আমি খুব খুশি	positive	সুন্দর খুশি	2	13
...	...	...	...	...	...
1136	ওই, কে রে শালা তুই?	negative	ওই রে শালা তুই	4	16
1137	সবকিছুই পরিকল্পনা মাফিক হবে	positive	সবকিছুই পরিকল্পনা মাফিক	3	23
1138	তুই কোন নরক থেকে এসেছিস রে?	negative	তুই নরক এসেছিস রে	4	18
1139	এখানেই মরবি তুই।	negative	মরবি তুই	2	9
1140	আপনি যা করেছেন তার জন্যে আপনাকে মেরে ফেলা উচিত	negative	জন্যে মেরে ফেলা	3	15

1141 rows × 5 columns

Figure 7. Sample of cleaned dataset

3.3. Machine learning algorithms and statistical analysis

About 571 records for positive and 570 records are for negative conversations in our dataset. For the dataset splitting purpose we used train-test split function. We followed supervised machine learning techniques. To train our model we used 80% of our data and for test 20% of data used. In number, 912 data used for trains and 229 data used for test purposes. To know the accuracy on our dataset we applied some classifier-based algorithms. These are support vector machine, multinomial naïve Bayes, k-nearest neighbors, logistic regression, decision tree, random forest, and stochastic gradient descent. In Figure 8, we have shown that how we have done our research shortly details.
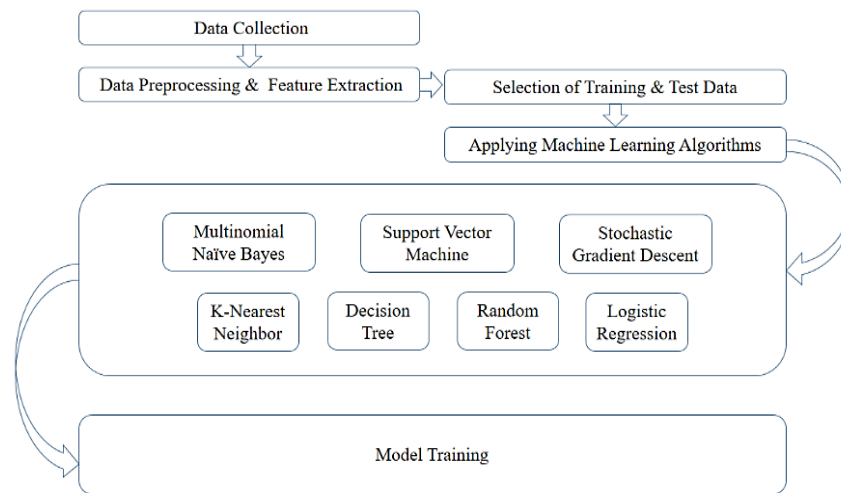
Figure 8. Proposed model structure

3.3.1. Feature extraction

We employ machine learning methods here to achieve natural language processing goals. Our model is trained by extracting all characteristics of each phrase from two primary characteristics. A method called tokenizer is presented here for this technique. Tokenizer divides phrases into words parts. These unique and common words have identical properties. In addition, TF-IDF is also such a numerical figure that examines the requirement of a term in a text. This approach is used by some important publications for several languages. Their success inspired us, and we found that our learning algorithms were the most accurate.

3.3.2. Classifier algorithms

It builds numerous decision trees during training. The naïve Bayes classification presupposes that there is no connection between the existence of a certain characteristic in a class and the presence of any other characteristic. This model is straightforward to create and beneficial for very big datasets in particular. Naïve Bayes even exceeds advanced categorization algorithms. The logical regression model may create a probability model from a class or event. To decide, for example, one group of images including photographs of different animals which may be investigated on a model of various classes. Stochastic gradient descent is renowned for improving any method transmitted particularly in machine learning algorithms in order to identify associated model parameters for both expected and actual results.

4. EXPERIMENTAL RESULT AND ANALYSIS

In this modern era, in intelligent analyzing of data and developing the related smart applications, the understanding of IoT [15]–[17], cyber-security [18], in particular, machine learning and deep learning [19]–[25] are crucial. According to our requirement, we update our model and dataset using machine learning approach. From this modification, we can accomplish that our used classifier is exactly usable for a wide range of use according to our dataset. As per our expectations, we achieved 86% accuracy from our proposed mode which is a fruitful outcome. This performance of the model creates a path to think about the improvement in results.

The research result was focused to identify whether a conversation is positive or negative. We have applied classifiers based on different machine learning models to extract the conversation type. The result has two criteria of positive and negative. There were 1141 data for training each of the models. We get various

accuracy on different models. Among 7 models the support vector machine and multinomial naive Bayes perform well with the highest accuracy. As we already discussed, we collect data from scripts as a conversation. All conversations have people's emotions like happy, sad, worried, annoyed, and motivated. We merged and categorized them into two main types, positive and negative. The decision-making capability of the classifiers was measured by their performance. Accuracy, precision, recall, and F-score were used to determine the performance of classifiers. For a classifier, the overall accuracy was considered an adequate standard. In the test set, it is necessary to have a notion of the correctly classified samples.

In Table 1 the accuracy scores obtained for the classifiers built are given. Here it is clear that the support vector machine gives the highest accuracy score of 0.85589 and multinomial naive Bayes gives almost similar accuracy of 0.8513. That is why it was needed to calculate the other performance measures to decide a suitable classifier for our dataset.

To measure the class agreement of the data labels with the positive labels given by the classifier the precision is used. We have to calculate the precision scores for each of the two-class labels because it is directly relevant to class labels. In Table 2 the values for each of the classifiers are given along with the 2 labels we used in this research work. We can see that the classifier random forest gives a score of 0.93 and multinomial naive Bayes gives 0.85 for positive conversation.

To identify class labels recall is known as sensitivity of the measurement that represents the effectiveness of the classifier. We also concentrated on achieving a score near 1 for the positive class label. The recall scores for two-class labels and classifiers are reported in Table 3. The decision tree and support vector machine had a recall score of 0.92 for positive dialogue. F1-score can be used to determine the relationship between positive labels and those provided by the classifier. The harmonic means of precision and recall for all two labels across all classifiers can be used to calculate it. The score close to 1 for the positive class label was considered when determining the optimum model of classifier. Table 4 shows the F1 scores for the class labels. Vector machines and multinomial naïve classifiers are supported by the classifiers. Bayes and stochastic gradient descent are the most effective methods for determining the best classifier for our dataset.

Table 1. Accuracy of classifiers	
Classifier	Accuracy
random forest	74.24%
decision tree	76.42%
logistic regression	82.53%
k-nearest neighbors	82.97%
stochastic gradient descent	83.41%
Multinomial naïve Bayes	85.15%
Support vector machine	85.59%

Table 2. Precision of classifiers	
Classifier	Precision
Random forest	67.01%
Decision tree	69.62%
Logistic regression	79.23%
K-nearest neighbors	79.39%
Stochastic gradient descent	79.55%
Multinomial naïve Bayes	85.96%
Support vector machine	81.68%

Table 3. Recall of classifiers	
Classifier	Recall
Random forest	96.55%
Decision tree	94.83%
Logistic regression	88.79%
K-nearest neighbors	89.66%
Stochastic gradient descent	90.52%
Multinomial naïve Bayes	84.48%
support vector machine	92.24%

Table 4. F1-score of classifiers	
Classifier	F1-Score
Random forest	79.15%
Decision tree	80.29%
Logistic regression	83.74%
K-nearest neighbors	84.21%
Stochastic gradient descent	84.68%
Multinomial naïve Bayes	85.22%
Support vector machine	86.64%

Our objective is to predict the mentally hampered individuals with higher precision which was achieved by random forest, multinomial naïve Bayes, and support vector machine. With remarkable accuracy support vector machine, multinomial naïve Bayes, and stochastic gradient descent perform well among the classifiers as shown in Table 5. Support vector machine, multinomial naive Bayes, and random forest all perform well as individual classifiers, as seen in the tables. Support vector machines work well for the challenge because our dataset is significantly more condensed, and the labels are poorly understood. K-nearest neighbor works effectively since there are fewer dimensions or attributes. The assumption of class conditional independence will only work for a large dataset, which is why the decision tree performs poorly in this case.

To avoid over fitting and robustness, it is needed to have a strong correlation over fitting nuts, though it is not exceptional. As it is not robust to noise and does not generalize well, future observed data decision trees do not work too well. In Figure 9 the overall performance comparison is shown.

Table 5. Performance analysis of different algorithms

Classifier	Accuracy	Precision	Recall	F1-Score
Random forest	74.24%	67.01%	96.55%	79.15%
Decision tree	76.42%	69.62%	94.83%	80.29%
Logistic regression	82.53%	79.23%	88.79%	83.74%
k-nearest neighbors	82.97%	79.39%	89.66%	84.21%
Stochastic gradient descent	83.41%	79.55%	90.52%	84.68%
Multinomial naïve Bayes	85.15%	85.96%	84.48%	85.22%
Support vector machine	85.59%	81.68%	92.24%	86.64%
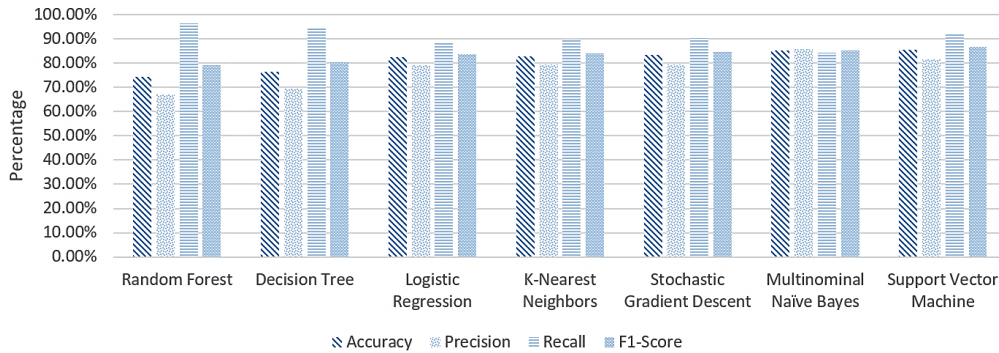
Figure 9. Performance analysis

4.1. Prediction

We have tried to test our model by using a random conversation data and we got a result. In Figures 10 and 11, We can see positive and negative prediction conversation. That Mean's, we can see that our proposed model can extract sentiment from Bangla conversation data.

```
model = open('cs_svm.pkl','rb')
svm_model = pickle.load(model)
Conversation = 'মামা কি অবস্থা আজকে তো এল ক্লাসিকো খেলা আছে কে জিতবে রিয়াল নাকি বার্সা তো খেলাটা কোন চ্যানেলে দেখাবে.... খেলা রাত 11 টায় সনি টিভিতে
processed_conversation = process_conversations(Conversation)
if (len(processed_conversation))>0:
    cv,feature_vector = calc_gram_tfidf(dataset.cleaned)
    feature = cv.transform([processed_conversation]).toarray()
    sentiment = svm_model.predict(feature)
    if (sentiment ==0):
        print(f"It is a Negative conversation")
    else:
        print(f"It is a Positive conversation")
else:
    print("This conversation doesn't contains any bengali Words, thus cannot predict the Sentiment.")

It is a Positive conversation
```

Figure 10. Predicting positive conversation

```
model = open('cs_svm.pkl','rb')
svm_model = pickle.load(model)
Conversation = 'কিরে তোর পড়াশোনার কী অবস্থা তুই নাকি পড়াশোনা করিস না সারাদিন খেলে বেড়ায় না ভাই আমি তো সারাদিনই পড়ি....'
processed_conversation = process_conversations(Conversation)
if (len(processed_conversation))>0:
    cv,feature_vector = calc_gram_tfidf(dataset.cleaned)
    feature = cv.transform([processed_conversation]).toarray()
    sentiment = svm_model.predict(feature)
    if (sentiment ==0):
        print(f"It is a Negative conversation")
    else:
        print(f"It is a Positive conversation")
else:
    print("This conversation doesn't contains any bengali Words, thus cannot predict the Sentiment.")

It is a Negative conversation
```

Figure 11. Predicting negative conversation

5.   CONCLUSION

This research work concludes with an expected outcome using machine learning approach of extracting sentiment from Bangla conversation data. Text mining and text analysis are very new terms in Bangla language. Though it is a tough task to work with some limitations, lacking the resources we tried to overcome these difficulties. Technology makes the communication sector easier with advancement. But embracing the advancement by ensuring the control of enormous data is necessary for us. We should be concerned about these terminologies to make the world of data more accessible and convenient.

6.   FUTURE WORK

This research work proposes a methodology that finds the scopes to work with Bangla conversation data. To accomplish that, machine learning models were trained from Bangla conversation data and able to extract sentiment from those conversations. There is a scope to apply a deep learning approach in our dataset to improve efficiency. Here in this work, we extract sentiment as a positive and negative category. But on a large scale, people's emotions, and sentiments as individuals like sadness, anger, neutral, happiness, and fear can also be extracted. For real-time conversation data, converting real-time conversations into text and analyzing sentiment from these conversations can also be done. However, scope lies in every possible opportunity. And opportunity revealed innovation and evolutions.

REFERENCES

[1]   C. O. Alm, D. Roth, and R. Sproat, "Emotions from text," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 579–586, doi: 10.3115/1220575.1220648.

[2]   C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proceeding of the 18th ACM conference on Information and knowledge management*, 2009, 375, doi: 10.1145/1645953.1646003.

[3]   T. Nasukawa and J. Yi, "Sentiment analysis: capturing favorability using natural language processing," in *Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003*, 2003, pp. 70–77, doi: 10.1145/945645.945658.

[4]   A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, vol. 1, pp. 142–150.

[5]   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, Mar. 2022, doi: 10.1016/j.array.2021.100123.

[6]   T. T. Aurpa, R. Sadik, and M. S. Ahmed, "Abusive Bangla comments detection on Facebook using transformer-based deep learning models," *Social Network Analysis and Mining*, vol. 12, no. 1, Dec. 2022, doi: 10.1007/s13278-021-00852-x.

[7]   M. R. H. K. Rahib, A. H. Tamim, M. Z. Tahmeed, and M. J. Hossain, "Emotion detection based on Bangladeshi people's social media response on COVID-19," *SN Computer Science*, vol. 3, no. 2, Mar. 2022, doi: 10.1007/s42979-022-01077-1.

[8]   M. A. Faruque, S. Rahman, P. Chakraborty, T. Choudhury, J.-S. Um, and T. P. Singh, "Ascertaining polarity of public opinions on Bangladesh cricket using machine learning techniques," *Spatial Information Research*, vol. 30, no. 1, pp. 1–8, Feb. 2022, doi: 10.1007/s41324-021-00403-8.

[9]   M. D. A. Iqbal, A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker, "BEmoC: a corpus for identifying emotion in Bengali texts," *SN Computer Science*, vol. 3, no. 2, Mar. 2022, doi: 10.1007/s42979-022-01028-w.

[10]   S. F. Shetu, M. Saifuzzaman, M. Parvin, N. N. Moon, R. Yousuf, and S. Sultana, "Identifying the writing style of bangla language using natural language processing," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225670.

[11]   M. M. R. Mamun, O. Sharif, and M. M. Hoque, "Classification of textual sentiment using ensemble technique," *SN Computer Science*, vol. 3, no. 1, Jan. 2022, doi: 10.1007/s42979-021-00922-z.

[12]   M. S. Neethu and R. Rajasree, "Sentiment analysis in Twitter using machine learning techniques," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2013, pp. 1–5, doi: 10.1109/ICCCNT.2013.6726818.

[13]   A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2016, pp. 628–632, doi: 10.1109/ICATCCT.2016.7912076.

[14]   M. Rahman and E. K. Dey, "Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation," *Data*, vol. 3, no. 2, May 2018, doi: 10.3390/data3020015.

[15]   M. Saifuzzaman, S. F. Shetu, N. N. Moon, F. N. Nur, and M. H. Ali, "IoT based street lighting using dual axis solar tracker and effective traffic management system using deep learning: bangladesh context," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICCCNT49239.2020.9225590.

[16]   M. Saifuzzaman, N. N. Moon, and F. N. Nur, "IoT based street lighting and traffic management system," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dec. 2017, pp. 121–124, doi: 10.1109/R10-HTC.2017.8288921.

[17]   R. Hasan, S. Islam, M. H. Rahman, M. Saifuzzaman, S. F. Shetu, and N. N. Moon, "Implementation of low cost real-time attendance management system: a comparative study," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Jun. 2020, pp. 1098–1101, doi: 10.1109/ICRITO48877.2020.9197764.

[18]   S. F. Shetu, M. Saifuzzaman, N. N. Moon, and F. N. Nur, "A survey of botnet in cyber security," in *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, Sep. 2019, pp. 174–177., doi: 10.1109/ICCT46177.2019.8969048.

[19]   K. K. Podder *et al.*, "Bangla sign language (BdSL) alphabets and numerals classification using a deep learning model," *Sensors*,

vol. 22, no. 2, Jan. 2022, doi: 10.3390/s22020574.

[20] M. Hossain *et al.*, "Prediction on domestic violence in Bangladesh during the COVID-19 outbreak using machine learning methods," *Applied System Innovation*, vol. 4, no. 4, Oct. 2021, doi: 10.3390/asi4040077.

[21] M. Al-Smadi, O. Qawasmeh, B. Talafha, and M. Quwaider, "Human annotated Arabic dataset of book reviews for aspect based sentiment analysis," in *2015 3rd International Conference on Future Internet of Things and Cloud*, Aug. 2015, pp. 726–730, doi: 10.1109/FiCloud.2015.62.

[22] L. Khan, A. Amjad, K. M. Afaq, and H.-T. Chang, "Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media," *Applied Sciences*, vol. 12, no. 5, Mar. 2022, doi: 10.3390/app12052694.

[23] M. Chen, K. Ubul, X. Xu, A. Aysa, and M. Muhammat, "Connecting text classification with image classification: a new preprocessing method for implicit sentiment text classification," *Sensors*, vol. 22, no. 5, Feb. 2022, doi: 10.3390/s22051899.

[24] K. Schouten, O. van der Weijde, F. Frasincar, and R. Dekker, "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data," *IEEE Transactions on Cybernetics*, vol. 48, no. 4, pp. 1263–1275, Apr. 2018, doi: 10.1109/TCYB.2017.2688801.

[25] H. Zou and K. Xiang, "Sentiment classification method based on blending of emoticons and short texts," *Entropy*, vol. 24, no. 3, Mar. 2022, doi: 10.3390/e24030398.

BIOGRAPHIES OF AUTHORS

Mahmudul Hassan ⓘ 🔍 SC Ⓟ studied Computer Science and Engineering from Daffodil International University. His main interests in research fields are natural language processing, image processing, machine learning and data mining. His research interests also include data science and computer vision. He can be contacted at email: mahmudul15-8991@diu.edu.bd.

Shahriar Shakil ⓘ 🔍 SC Ⓟ studied Computer Science and Engineering from Daffodil International University. His main interests in research fields are image processing, machine learning and data mining. His research interests also include data science and computer vision. He can be contacted at email: shahriar15-8558@diu.edu.bd.

Nazmun Nessa Moon ⓘ 🔍 SC Ⓟ is an associate professor of the Department of Computer Science and Engineering at Daffodil International University. She received the B.Sc. degree in Computer Science and Engineering from Rajshahi University of Engineering and Technology and M.Sc. in Information and Communication Technology from Bangladesh University of Engineering and Technology (BUET). Her interested research fields are IoT, digital image processing and machine learning. She can be contacted at email: moon@daffodilvarsity.edu.bd.

Mohammad Monirul Islam ⓘ 🔍 SC Ⓟ is a lecturer (senior scale) of the Department of Computer Science and Engineering at Daffodil International University. He completed his B.Sc. (India) in Computer Science and M.Sc. (UK) in Computing. His areas of interests are navigation for organizations/institutions and data warehousing. He can be contacted at email: monirul@daffodilvarsity.edu.bd.

Refath Ara Hossain 🆔 8ª SC Ⓟ is a lecturer of the Department of Computer Science and Engineering at Daffodil International University. Her interested research fields are data mining, digital image processing and machine learning. She can be contacted at email: refath.cse@diu.edu.bd.

Asma Mariam 🆔 8ª SC Ⓟ is a lecturer of the Department of Computer Science and Engineering at Daffodil International University. Her interested research fields are digital image processing, data mining and machine learning. She can be contacted at email: asma.cse@fiu.edu.bd.

Fernaz Narin Nur 🆔 8ª SC Ⓟ is an associate professor at the Notre Dame University Bangladesh (NDUB) in the Department of CSE. She is a passionate researcher in the fields of wireless sensor network, cloud computing, internet of things, and performance analysis. She can be contacted at email: fernazcse@ndub.edu.bd.