Analysis of the Opinion of the People of Bangladesh on the Padma Setu Megaproject

Tamim Al Mahmud
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

Sazeda Sultana
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

Tanjin Irfan Chowdhury
Department of Computer Science and Engineering
Green University of Bangladesh
Dhaka, Bangladesh

ABSTRACT

Sentiment analysis is the term used to describe the process of mining people's opinions or emotions. The public opinion on Padma Bridge was researched and documented in this research paper. Considering the results, the Bangladeshi government can easily decide on future large construction projects. The construction of the Padma bridge is an important milestone for Bangladesh because to the high budget and prohibited World Bank funding. On YouTube, Facebook and other social online news, Bangladeshis share their thoughts, feelings, suggestions, and opinions about the Padma Bridge project. The main objective of the research is sentiment analysis of Padma Bridge sentiment based on the Bangla comment dataset. We collected over 10,000 data with two categories of sentiment: positive and negative. The innovative voting method proposed in this paper is a significant breakthrough in the field of sentiment analysis. Our approach can compare and count the sentiments generated by different Machine Learning and Deep Learning models, resulting in a decision-making process that is more accurate and consistent. Our model considers the strengths and weaknesses of each individual model, ensuring that the final decision is based on the maximum voting results. Our research shows that approach outperforms every machine learning and deep learning model by about 6.5% in terms of accuracy. This improvement is significant and has practical implications for industries such as marketing, finance, and politics, where accurate sentiment analysis is crucial for decision-making. Our approach has the potential to revolutionize sentiment analysis by providing a more robust and accurate method for analyzing large volumes of data. Further research could explore ways to optimize this approach even further, making it even more effective in real world applications.

General Terms

Sentiment Analysis, Machine Learning, Deep Learning

Keywords

Sentiment, SVM, RF, LSTM, Word Cloud, Pre-Processing, BoW, SVD.

1. INTRODUCTION

Sentiment analysis describes an individual's viewpoint, feeling, or response to a subject. Different situations call for using the concept of sentiment analysis. Establishing how individuals respond to other things can likewise be done using sentiment analysis. For instance, a company may conduct an online survey to find out how consumers feel about any of its most recent items and then use sentiment analysis to figure out how well-received they are. As a result, the business can decide whether it needs to change or improve its products. Now that everything can be easily gathered, we are in the modern age of the internet. Due to the Internet's ability to connect individuals from all over the world, data for sentiment analysis is typically gathered from surveys or a variety of sources, including social media, online publications, and news portals. The internet era has completely changed how we get data. It's simpler than ever to collect data for sentiment analysis because the globe is so interconnected. Just two of the various avenues that may be used to get information about people's ideas and feelings are surveys and social media. Internet-based publications like magazines and news sites are a great source of information on public opinion on various subjects. This vast amount of data allows us to analyze trends and patterns in human behavior, which can be incredibly useful for businesses, governments, and organizations looking to understand their audience better.

However, with great power comes great responsibility, and it's essential to ensure that this data is collected ethically and used in ways that benefit society. As we continue to navigate the age of the internet, it's crucial that we use data collection methods responsibly to create a more informed and connected world. In order to improve the service and the product, it is essential to analyze consumers impressions. We decided to concentrate on Bengali, which is spoken by 300 million people as their primary and secondary language, respectively. However, there has not been much study done on Bangla. To enhance the research on Bengali, we are using sentiment analysis on this language. While this language is simple and easy to figure out for people, it is highly challenging for machines or computers to understand. Making the machine understand Bengali is therefore our main problem.

The Padma Bridge has been recognized as one of Bangladesh's most significant construction initiatives. People's perspectives on the bridge vary since Bengali people are directly associated with it. They share their own perspectives with a variety of sources online, expressing their thoughts and ideas in various ways. Therefore, the number of people has a significant impact on the government's decision to move forward with new, major initiatives. In this research, we applied machine learning models such as Support Vector Machine (SVM), Random Forest (RF) and Linear Support Vector Classifier (LSVC), as well as one deep learning model which called Long Short-Term Memory (LSTM). We collected our data from YouTube, internet news sites, and Facebook comments.

The remaining part of the paper is structured as follows: The next sections are related works in section 2, methodology in section 3, performance evaluation in section 4, conclusion in section 5 and reference in section 6.

2. RELATED WROK

Many different methods of language have given considerable. attention to sentiment analysis in recent years. Numerous languages have participated in similar research projects. We have collected and analyzed a variety of recently published research articles on sentiment analysis in several languages for our study. This study was conducted using a variety of languages, including Bengali, English, Arabic, Spanish, Hindi, Tamil, and Urdu. On datasets of many languages, researchers have performed sentiment analysis using various machine learning models and deep learning methods.

2.1 Work In Bangla

In this study, the authors applied a Random Forest Classifier to evaluate the total positivity and negativity of a document or sentence utilizing unigram, part-of-speech (POS)labeling, and negation handling. Their proposed model improved by 87% [1]. In this paper, they provide a word embedding approach for provide a word embedding approach for each word vectorization and use long-short-term memory (LSTM) to deal with long-term dependencies. They also created word embedding using the word2vec technique, which creates a matrix of values from a collection of text. Spelling checking and stemming aren't included in the initial processing step of the gathered dataset, which is a drawback of this proposal model [2]. In this study, the authors proposed the use of three different text categorization models: long-short-term memory (neural network), convolutional neural networks (CNN), and a binary support vector machine classifier. These models were used for a baseline assessment to verify the accuracy of the data set. With 74.74 percent, they reach the LSTM models. highest accuracy [3]. Articles [4]-[8] worked on the Bengali. language different areas.

2.2 Work In English

In this paper, text and emojis were used for multi modal sentiment analysis. A dataset comprising evaluations of the Samsung M21 Mobile across different social media platforms was prepared for this. They gathered a total of 9003 points of data. There was a total of 6519, or 74%, positive data and 2326, or 26%, negative data. A CSV file comprising text and emoji data has been added to the dataset. The initial processing of the dataset involved six processes, including lowercase conversion, stop words, word tokenization, lemmatization, drop number, and punctuation. Used Word2Vec and Skip Gram with CBOW as two feature extraction techniques. The algorithm made use of LSTM, CNN, CNN-LSTM, and Bi-LSTM. They could create their recommended model without the use of emoji filters [9]. In this study, the most productive machine learning method for sentiment analysis in Taglish, English, and Filipino is determined. A dataset on mental health emergencies is being created by gathering tweets. 5385 pieces of information have been collected, of which 3085 are positive and 2310 are negative. They are employing four machine learning techniques: naive Bayes, stochastic gradient descent classifier, logistic regression, and linear support vector classifier. Furthermore, they employ feature extraction to provide counts of frequencies, add punctuation, mark sections, remove words to avoid, and tokenize words. Their logistic regression model accuracy of 81% is the greatest achievement in machine learning [10]. The authors of this study examined perceptions about the categories of business reviews. They took advantage of a dataset containing 1.6 million records that "Yelp" offered. They employed two feature extraction techniques as well as the linear support vector classification, logistic regression, naive bayes, and stochastic gradient descent classifier machine learning models, which together included four different machine learning models. 80% for training and 20% for testing should be applied after splitting the dataset. In order to prepare the data, they utilized the Natural Language Toolkit (NLTK) [11]. They recommended an article that uses a convolution network and a bidirectional long-short-term memory to generate a dataset of online troll reviewers that was gathered from Reddit social media. Additionally, this data relates to political choices made online. They created a dataset with 6695 trolling review collections and 10,000 no-troller data points. Text and numerical data were used in two distinct 2 trials. There are 12 attributes in this dataset. The data has gone through pre-processing to make it anonymous and unrestricted. The data was divided so that 20% went to the testing set, 10% went to the validation set, and 80% went to the training set. The word embedding layer has a 20,000-word vocabulary, a 50-word embedding dimension and an acceptable data length of 4 to 150 boards. For the outcomes of the experiment, they employed accuracy, fi-score, sensitivity, and precision. 100 percent accuracy for experiment-based numerical data and 97 percent accuracy for experiment-based text data [12]. The authors of this paper used Twitter tweets to do sentiment analysis. Aside from it, several methods and techniques for sentiment analysis were covered it, several methods and techniques for sentiment analysis were covered. They gathered Twitter tweets and turned them into a dataset with three classifications, including positive, neural, and negative. 70 percent of the dataset is used for training tests, while 30 percent is used for testing tests. Five algorithms—KNN, Extra Tree Classifier, Naive Bayes, Lexicon-Based, and Support Vector Machine—are used to analyze this shared training and testing data. Accurate measurements of the outcome were made using recall, fi-score, precision, and accuracy. The maximum accuracy, which abstained by the pipeline with Naive Based, was 79% [13].

2.3 Work In Arabic

In this study, they explored the sarcasm of the Arabic language using the sentiment-based dataset "ArSarcasm" in Arabic. They applied a deep learning model called BiLSTM or sarcasm detection. Sarcastic and non-sarcastic data were identified in the dataset, respectively. Therefore, they separated them into three groups: positive, negative, and neutral. Their BiLSTM model has a remarkable 46% F1 score [14]. Sentiment analysis of newspaper comments from Algeria was conducted for this article. From the internet newspaper in Algeria, they gathered the newspaper comments. Three websites: Enahar, Elkhabar, and Echprouk were subsequently pursued. There are articles on topics connected to politics, sports, news, and religion. The dataset was annotated based on adjudicator and annotation criteria. After performing operations like stop word removal, monotonous letter removal, tokenization, word vector, human preprocessing of text, stemming, and word n-gram on the full data, Employing the three extraction techniques of unigram, trigram, and bi-gram, all three algorithms are applied sequentially. way KNN, SVM, and NB worked Without the light steaming of the dataset, the NB trigram had the maximum accuracy of 89.20 percent, whereas the NB bigram had the highest accuracy of 89.20 percent [15].

3. METHODOLOGY

This section will go over our working method and the models we applied. We employed the Python programming language and its corresponding tools throughout the project. Figure 1: The architecture of our concept.

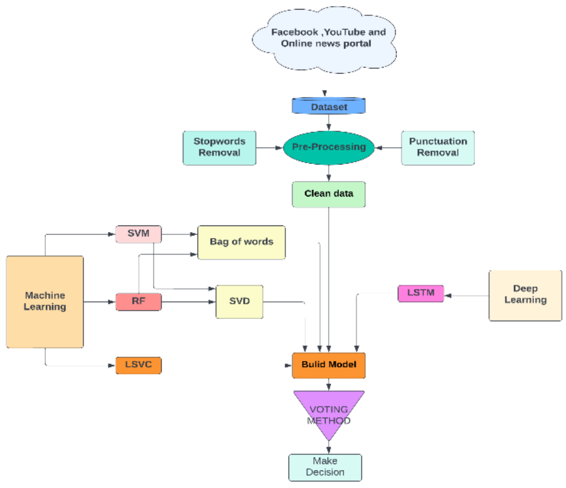

Figure 1: Our Architecture Concept

3.1 Dataset

Data is commonly referred to as the brain of artificial intelligence. In comparison to the data, we can provide. The machine can evaluate this data more accurately. For our model, we created a dataset of 10,000 data points. Serial, comment, sentiment, and source are the four columns in our dataset. Figure 3.1: Display our Dataset. We collected the data from several reliable sources. While we were researching public opinion on Padma Bridge, we obtained public opinion on Padma Bridge from several sources. In Figure 3.2, a pie chart illustrating several sources is shown.

	Serial	comment	sentiment	Source
0	0	আমার কাছে আনন্দ লাগতেছে খবরটা শুনে পদ্মাসেতু চ...	Positive	YouTube
1	1	2030 সাল নাগাদ চালু	Positive	YouTube
2	2	ব্যক্তিত্ব আর জেদ থাকলে যেকোনো কিছু অর্জন সম্ভ...	Positive	YouTube
3	3	সেতুটা কাছে থেকে দেখে স্যালুট করেছিলাম আমাদের ...	Positive	YouTube
4	4	সুন্দর ভাবে সেতুটি তৈরি করার জন্য মাননীয় প্রধা...	Positive	YouTube
...	...	...	...	...
9996	9996	মাননীয় প্রধানমন্ত্রী কে অসংখ্য ধন্যবাদ জানিয়ে ...	Positive	YouTube
9997	9997	আমরা দীর্ঘদিন ধরে যা স্বপ্ন দেখছিলাম তা অবশেষে...	Positive	YouTube
9998	9998	অর্থনীতি ও জীবনযাত্রার মানের উন্নয়নের লক্ষে এ...	Positive	YouTube
9999	9999	পদ্মা সেতু বাংলাদেশে অনেক মানুষের জীবন পরিবর্...	Positive	YouTube
10000	10000	এগিয়ে যাও বাংলাদেশ। ১৯৭১ সালের স্বাধীনতার পরে ...	Positive	YouTube

Figure 2: Our Padma Bridge Dataset

The sources from which we gathered data includes Padma Bridge-related Facebook postings, YouTube comments, and comments from significant Bengali online news portals (Prothom Alo, Jamuna TV, Manab Zamin, etc.). where individuals have left comments about the Padma Bridge. From our sources, we have simply gathered useful comments in Bangla. There are sufficient reasons for collecting data from these sources; typically, Facebook or YouTube are very appreciated websites that users can use easily.

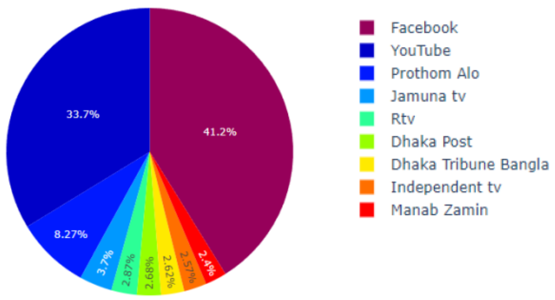

Figure 3: Pie chart of Several Sources

Additionally, the fact that there are so few comments currently that are authentic and so few people simply comment on the subject makes our internet-based newspaper site a great way for collecting data. Facebook and YouTube provide fewer benefits. Because there are many comments, there are also some that are unsubstantiated or off-topic. 10,000 points of data represented the dataset we created. Serial, comment, sentiment, and source are the four columns that make up our dataset. Positive and Negative values were separated from the dataset. A pie chart containing the percentages of the various emotions is shown in Figure 3.4.

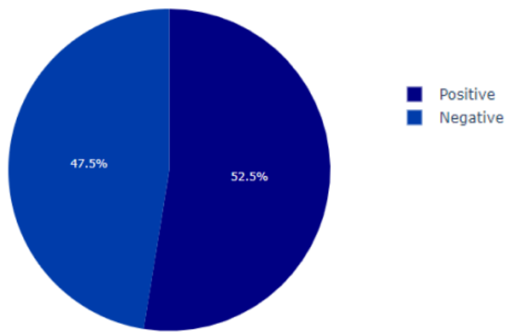

Figure 4: Pie chart of Several Sentiment

Based on positive and negative emotions, we displayed the text length of our complete dataset. The data describes how our dataset's count, mean, standard deviation, mini- mum, and maximum values function. By counting its data, the function "Count" displays the total number of comments. The average of the data is displayed as the mean; the mean is the shortest length of the comment; the maximum is the longest length of the comment in the whole positive data; and the std is the standard deviation that describes how the values are distributed. Positive sentiment has 5248 points. The average length of comments with positive sentiment is 15.24, with the smallest length being 2.0 and the largest being 70.0. The standard deviation is 8.17, and 25% of the dataset's lengths are 9.0, 50% are 13.0, and 75% are 19.0. The number of negative emotions is 4741. The average length of comments with negative sentiment is 15.94, with the minimum and greatest lengths of 2.0 and 89.0, respectively, and an STD of 8.33. The dataset's 25% length is 10.0, 50% length is 14.0, and 75% length is 20.0.A positive text length distribution is shown in Figure 3.5, and a negative text length distribution is shown in Figure 3.6.

3.2 Word Cloud

A word cloud is a text comprehension tool that dynamically displays text using various hues to represent the most frequently used words. Word clouds are also known by the names tag cloud and text cloud. From our dataset's comments,

	length
count	5260.0
mean	15.2
std	8.17
min	2.0
25%	9.0
50%	13.0
75%	19.0
max	70.0

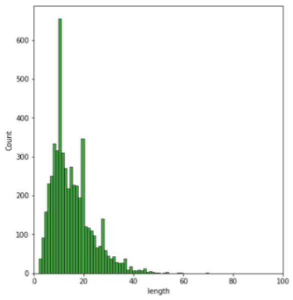

Figure 5: Text Length Distribution for Positive Sentiment

	length
count	4741.0
mean	15.94
std	8.33
min	2.0
25%	10.0
50%	14.0
75%	20.0
max	89.0

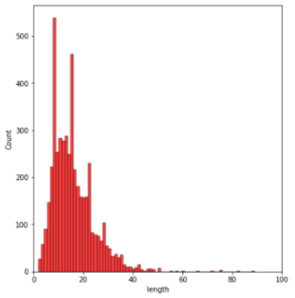

Figure 6: Text Length Distribution for Negative Sentiment

we generated two-word clouds. While the number of low-frequency words is comparatively small, the size of the terms in the dataset with high frequency in the cloud is significant. Positive and negative word clouds are shown in Figs. 3.7 and 3.8, respectively, as are word clouds created from comments. Despite our best efforts, we were not able to correct Word’s conjugation letters.

Figure 7: Word Cloud of Positive Sentiment

Figure 8: Word Cloud of Negative Sentiment

3.3 Pre-processing

Pre-Processing: Data processing is essential for data analysis. The opportunity of correct output without data processing is quite low. Therefore, we cleaned the data using the following procedures.

(1) Punctuation Removal: Cleaning up a dataset is a crucial step in data analysis and removing extra punctuation marks is one of the essential tasks in this process [16]. These marks include: ,, :, ;, ”, ’, ?,<, >, -, &, @, #, !, *, , , $, %, ... etc. These punctuation marks can create inconsistencies in the dataset and affect the accuracy of the analysis. Removing them ensures that the data is consistent and reliable. However, it’s important to note that some punctuation marks may have specific meanings or uses in certain contexts. Therefore, it’s essential to carefully review the data before removing any punctuation. Once the extra punctuation has been re moved from the dataset, it can be further processed and analyzed with confidence to extract meaningful insights that can help drive inform decision.

(2) Stop Word Removal: In sentiment analysis, stop words can skew the results and make it harder to accurately classify the sentiment of a piece of text. Removing stop words helps reduce noise and improve the accuracy of sentiment analysis models. However, it is important to note that not all stop words should be removed, as some may carry important contextual information. It is up to the data scientist or researcher to determine which stop words are necessary for the specific task at hand. Words can be: "অথব", "আছ", "আমর", "ইত", "ইহ", "এক","এতটা", "এমনক", "করছ", "করত", "করব", "করল", "কয়", "চল","জনক", "তথ", "তব", "তর", "নও", "নজন", "নত", "ফল", "বক", "বদল", "বল", "বলত", "বস", "বহ", "মত", "মধ", "মন", "যতট","যথ", "যদ", "যম", "রও", "রণ", "রত", "হইত", "হইব" etc.

(3) Drop Numbers: Excluding all types of numbers from a dataset can have both advantages and disadvantages. On the one hand, it can help simplify the analysis process by removing any numerical variables that may complicate the interpretation of results. This can be particularly useful when dealing with qualitative data or when looking for patterns and trends that are not necessarily tied to numerical values such as 1, 2, 3, 4, 5, 6, 7, 8, 9,0, "০", "১", "২", "৩", "৪", "৫", "৬", "৭", "৮", "৯" etc.

3.4 Training Data and Testing Data

The dataset has been separated into two fundamental categories: training and testing. The training set is used to train the machine learning model and deep learning model, while the testing set is used to evaluate its performance. Training data representing 80% of the dataset and test data for 20%. For reliable and efficient artificial intelligence models that can generalize well to new data, the dataset must be divided and represented in the proper way.

3.5 Feature Extraction

Text data is a crucial component of machine learning models, and understanding how to model and test it is essential. One popular method for text data modeling is the “bag of words” approach, which involves representing text as a collection of words without considering their order or context. This method makes it simple to compare and analyze text data. The Singular Value Decomposition (SVD) approach is an effective methodology that breaks down a matrix into its constituent elements and identifies fundamental patterns and associations that exist within the given data. The singular value decomposition (SVD) can be utilized to lower the dimensionality of textual data, thus rendering it more amenable to analysis and processing. Through the integration of these techniques, machine learning models can proficiently analyze vast quantities of textual data, thereby enhancing their precision and efficacy. While these approaches are effective for modeling and evaluating text data, it’s crucial to keep in mind that they have some drawbacks and should be combined with other methods for the best outcomes.

3.6 Approach

In the domain of data science, Support Vector Machine, Random Forest, and Linear Support Vector Classifier are prominent machine learning models that are widely utilized.

(1) Support Vector Machine (SVM): The support vector machine (SVM) classifier is a well-liked technique that is especially helpful when there are just two classifications in the dataset. An SVM can do classification by splitting the two sentiment datasets using the best hyperplane to separate all the datasets into one class from those of the others. SVMs can, however, also be used to solve regression issues by identifying a hyper-plane that minimizes error while best fitting the data points.

(2) Random Forest (RF): The implementation of the Random Forest approach is widely recognized as a prominent method for achieving this objective. During the training process, this methodology involves the construction of decision trees that subsequently produce sentiments reflecting the classification tree approach. The process of constructing decision trees involves the random selection of subsets of features and data points, which serves to alleviate the problem of over-fitting and enhance the accuracy of the resulting tree model. The sentiment that is produced has the potential to facilitate the categorization of novel data points into distinct categories based on their resemblance to the data that has been used for training purposes. The Random Forest algorithm is a highly advantageous approach for managing extensive datasets, especially when intricate correlations exist among variables. Notably, this algorithm can adeptly handle both categorical and continuous data types.

(3) Linear Support Vector Classification (LSVC): When working with datasets that can be separated into two groups using only a hyperplane, it is especially useful. As a result, it may be used to group data points according to how they are distributed along a single line or plane. The method operates by identifying the best hyper-plane or line that, with the greatest possible mar- gin, divides the two classes of data. By doing this, the classifier is made to be accurate and reliable even while working with noisy or complicated datasets. In deep learning, we use only the Long Short-Term Memory type of method.

(1) Long Short-Term Memory (LSTM): The field of deep learning has been significantly transformed by the Long Short-Term Memory (LSTM) algorithm, a robust and dominant technique. The recurrent neural network belongs to a class of computational models that is specifically tailored to process sequential data and has demonstrated exemplary performance in various domains, including but not limited to speech recognition, natural language processing, and time series analysis. The Long Short-Term Memory (LSTM) is known for its capability to retain and incorporate information relating to extensive dependencies within the data, thereby rendering it well-suited for the purpose of constructing models for intricate sequences. One additional benefit is its inherent flexibility, as it can be tailored to accommodate diverse data sets and a variety of tasks.

3.7 Algorithm of Our Proposed Voting System

Each comment on our platform is subject to analysis through our voting system algorithm. This algorithm utilizes four different algorithms to check the data and predict whether a comment is positive or negative. If there are more positive votes, the comment is recognized as positive, and if there are more negative votes, it is recognized as negative. Our algorithm has been designed to achieve the highest accuracy possible, ensuring that our users can trust the feedback they receive on their comments. By utilizing multiple algorithms and analyzing various data points, we can provide a com-prehensive analysis of each comment. Our proposed algorithm has been thoroughly tested and illustrated below, demonstrating its effectiveness in accurately predicting the sentiment of each comment. With this system in place, we can maintain a positive and productive community on our platform while providing valuable feedback to our users.

Algorithm 1 Algorithm for Our Proposed Voting System

Input: Comment (Com)

Output: S = (pos, neg),

Here S = Sentiment, pos=Positive and, neg=Negative

Initial: $Sum_{pos} = 0$ and $Sum_{neg} = 0$

For every sentiment Seni in results

if Sen_i is Positive then

$Sum_{pos} \leftarrow Sum_{pos} + 1$

else if Sen_i is Negative then

$Sum_{neg} \leftarrow Sum_{neg} + 1$

end if

For every comment Com_i

if $Sum_{pos} > Sum_{neg}$ then

$Com_i \leftarrow$ Positive

else if $Sum_{neg} > Sum_{pos}$ then

$Com_i \leftarrow$ Negative

end if

4. PERFORMANCE EVULATION

Performance evaluation is used to measure every search. Here, several Machine Learning and Deep Learning model's performance has been evaluated employing various kinds of metrics include Precision, Recall, Accuracy, and F-measure.

Precision- A statistical parameter known as precision evaluates the accuracy of positive predictions made by classification models. Based on all the positive predictions made by the model, it calculates the percentage of true positive predictions (positive instances that were correctly categorized).

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

Recall- A statistical parameter called recall, frequently referred to as sensitivity or true positive rate, measures how well a classification model can detect positive examples. Out of all actual positive occurrences in the dataset, it calculates the percentage of true positive predictions (positive instances that were correctly categorized). The model is effective in identifying the positive occurrence in the data when recollecting is high.

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

Accuracy- A statistical parameter called accuracy evaluates how accurate a classification model's predictions are on average. It measures the percentage of the dataset's total number of instances that were correctly classified, including both true positives and true negatives. Although accuracy offers a broad evaluation of the model's performance, it may not be suitable for datasets with imbalanced classes.

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

F-measure- A statistical metric called the F-measure, commonly referred to as the F1-score, is used to assess how well categorization models perform. It provides an accurate measure of a model's accuracy by combining precision and recall into a single statistic. It considers the model's accuracy (ability to identify positive examples properly) and recall (ability to locate all pertinent positive instances).

$$\text{F-measure} = 2 * \frac{P\ recision \times Recall}{(P\ recision + Recall)}$$

The terms can be described using alternative terminology presented in the following words:
(1) True Positive (TP): The instances that a classification model accurately classifies as positive are known as True Positives, or TP. To phrase it another way, it is the quantity of positive cases in the dataset that the model correctly identifies as positive.
2) True Negative (TN): The instances that a classification model correctly classifies as negative are known as True Negative, or TN. In other words, it's the number of instances in the dataset that the prediction was correctly identified as negative.
(3) False Positive (FP): A classification model may inaccurately classify some cases as positive, which is known as a false positive (FP). It is, in other words, the number of positive instances in the dataset that the model incorrectly identifies as positive.
(4) False Negative (FN): False Negative, also known as FN, is the term used to describe situations where a classification model incorrectly classifies them as negative.
In other words, it refers to how many instances of negativity in the data set the model incorrectly categorized as negative. The algorithm was applied to the dataset, and the training outcomes are displayed in the below table. Table 4.1 shows that we measured performance using accuracy, precision, recall, and f-measure.

Utilizing feature extraction from the Bag of Words, we achieved the best level of accuracy in the Support Vector Machine. The Bag of Words dataset has a well-defined structure that the machine can understand. We have therefore achieved the highest level of accuracy. Singular Value De-composition, which splits the dataset into a matrix utilizing SVM to extract features, was also used. It merely decreases 42 in length of data but does not provide a strong structure like Bag of Words, leading to lower accuracy than BoW. However, as a feature of the Random Forest method, we used Bag of Words and Singular Value Decomposition. Because the dataset's data is separated into trees in the random forest algorithm, understanding the machine takes a longer time, and the algorithm's accuracy is lower than that of the SVM algorithm. As a result, we used the function of sigmoid activation in linear short-term memory, which has two layers, in deep learning.

Table 1: Results of the dataset's training for SVM, Random Forest, Linear SVC, and LSTM

The Outcomes of Training						
Statistic Measure	**SVM with BoW**	**SVM with SVD**	**RF with BoW**	**RF with SVD**	**Linear -SVC**	**LSTM**
Precision	82.66 %	80.63 %	80.96 %	78.31 %	80.57 %	76.47 % (Phase 1/4)
Recall	82.69 %	80.61 %	80.03 %	78.37 %	80.65 %	77.18 % (Phase 2/4)
F-Measure	82.45 %	80.33 %	80.86 %	78.32 %	80.50 %	77.23 % (Phase 3/4)
Accuracy	82.45 %	80.33 %	80.87 %	78.38 %	80.51 %	77.26 % (Phase 4/4)
Algorithm for Our Proposed Method Result is: 88.25%						

5. CONCLUSION

The Padma Bridge has been a topic of debate among people for quite some time now. It is no surprise that differing perspectives about it are common. While some believe that the bridge will bring development and prosperity to the region, others are skeptical about its impact on the environment and local communities. To determine the opinions of the public, we have gathered their ideas about Padma Bridge through various channels. We utilized deep learning and machine learning methods to analyze our datasets. While many people are enthusiastic about the potential advantages of the bridge, our investigation showed that there are also improving areas regarding its building process and long-term effects on the mega-project. The study's findings suggest that most people have positive opinions about the significant initiative in Bangladesh. This is a positive sign for the government, as it suggests that there is public support for such mega projects. The data also suggests that there is room for further development in this area, and that the government could pursue other similar projects in the future. This could have significant implications for the country's economy and infrastructure, as well as its overall development. Furthermore, the study's findings are likely to be of interest to future researchers who are studying Bengali texts, as they provide valuable insights into public opinion and attitudes towards large-scale projects in Bangladesh. Overall, this research has important implications for both policymakers and scholars alike and highlights the need for continued investment in infrastructure and development in Bangladesh.

6. ACKNOWLEDGMENTS

Our thanks to the reviewers for their scholarly review and help to enhance the quality of the article. Furthermore, thanks to the experts who have contributed towards development of the template.

7. REFERENCES

[1] N. Tabassum and M. I. Khan, "Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-5, https://doi.org/10.1109/ECACE.2019.8679347

[2] M. F. Wahid, M. J. Hasan and M. S. Alom, "Cricket Sentiment Analysis from Bangla Text Using Recurrent Neural Network with Long Short Term Memory Model," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4, https://doi.org/10.1109/ICBSLP47725.2019.201500

[3] M. A. -U. -Z. Ashik, S. Shovon and S. Haque, "Data Set For Sentiment Analysis On Bengali News Comments And Its Baseline Evaluation," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-5, https://doi.org/10.1109/ICBSLP47725.2019.201497

[4] T. A. Mahmud, S. Sultana, T. I. Chowdhury and F.R. Anando, "A New Approach to Analysis of Public Sentiment on Padma Bridge in Bangla Text," 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022, pp. 1-6, https://doi.org/10.1109/STI56238.2022.10103315

[5] M. G. Hussain, T. A. Mahmud and W. Akthar, "An Approach to Detect Abusive Bangla Text," 2018 International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 2018, pp. 1-5, https://doi.org/10.1109/CIET.2018.8660863

[6] M. G. Hussain, S. Kabir, T. A. Mahmud, A. Khatun and M. J. Islam, "Assessment of Bangla Descriptive Answer Script Digitally," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4, https://doi.org/10.1109/ICBSLP47725.2019.202042

[7] T. Ahmed, S. F. Mukta, T. Al Mahmud, S. A. Hasan and M. Gulzar Hussain, "Bangla Text Emotion Classification using LR, MNB and MLP with TF-IDF CountVectorizer," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 275-280, https://doi.org/10.1109/ICSEC56337.2022.10049341

[8] Md Gulzar Hussain, Tamim Al Mahmud. (2019). A Technique for Perceiving Abusive Bangla Comments. GREEN UNIVERSITY OF BANGLADESH JOURNAL OF SCIENCE AND ENGINEERING, 04(01). https://doi.org/10.5281/zenodo.3544583

[9] Kumar, T. Praveen, and B. Vishnu Vardhan. "A Pragmatic Approach to Emoji based Multimodal Sentiment Analysis using Deep Neural Networks." Journal of Algebraic Statistics 13.1 (2022): 473-482. https://doi.org/10.52783/jas.v13i1.108

[10] Aliman, G.B., Nivera, T.F.S., Olazo, J.C.A., Ramos, D.J.P., Sanchez, C.D.B., Amado, T.M., Arago, N.M., Jorda Jr, R.L., Virrey, G.C. and Valenzuela, I.C., 2022. Sentiment Analysis using Logistic Regression. Journal of Computational Innovations and Engineering Applications JULY 2022, 35, p.40. https://www.dlsu.edu.ph/wp-content/uploads/pdf/research/journals/jciea/vol-7-1/4aliman.pdf

[11] Salinca, Andreea. (2015). Business Reviews Classification Using Sentiment Analysis. 247-250. https://doi.org/10.1109/SYNASC.2015.46.

[12] Al-Adhaileh, Mosleh & Aldhyani, Theyazn & Alghamdi, Ans. (2022). Online Troll Reviewer Detection Using Deep Learning Techniques. Applied Bionics and Biomechanics. 2022. 1-10. https://doi.org/10.1155/2022/4637594

[13] Iram, M. ., Rehman, S. U., Shahid, S. ., & Mehmood , S. A. . (2022). Anatomy of Sentiment Analysis of Tweets Using Machine Learning Approach : Anatomy of Sentiment Analysis of Tweets. Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences, 59(2), 61–73. https://doi.org/10.53560/PPASA(59-2)771

[14] Ibrahim Abu Farha and Walid Magdy. 2020. From Arabic Sentiment Analysis to Sarcasm Detection: The ArSarcasm Dataset. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, pages 32–39, Marseille, France. European Language Resource Association. https://aclanthology.org/2020.osact-1.5

[15] Hichem Rahab, Abdelhafid Zitouni, and Mahieddine Djoudi. 2021. SANA: Sentiment analysis on newspapers comments in Algeria. J. King Saud Univ. Comput. Inf. Sci. 33, 7 (Sep 2021), 899–907. https://doi.org/10.1016/j.jksuci.2019.04.012

[16] Tamim Al Mahmud, Md Gulzar Hussain, Sumaiya Kabir, Hasnain Ahmad, and Mahmudus Sobhan. 2020. A Keyword Based Technique to Evaluate Broad Question Answer Script. In Proceedings of the 2020 9th International Conference on Software and Computer Applications (ICSCA 2020). Association for Computing Machinery, New York, NY, USA, 167–171. https://doi.org/10.1145/3384544.33846048