# BanglaHeartbeat: Mapping Emotional Echoes in Social Media Texts

Bidyarthi Paul [1], SM Musfiqur Rahman [2],Md. Ziaul Hasan[3],Dipta Biswas [4]
[1,2,3,4]Department of Computer Science and Engineering,
Ahsanullah University of Science and Technology, Dhaka, Bangladesh

***Abstract*—Understanding emotions in written language is a growing area of research, especially for languages like Bangla that aren't widely studied and have their own local expressions and cultural details. This study explores emotion analysis in the EmoNoBa dataset, focusing on 22,698 social media comments. We use machine learning models—Linear SVM, KNN, and Random Forest—with n-gram features from a TF-IDF vectorizer for linguistic analysis. We also examine the impact of PCA for dimensionality reduction. Additionally, we enhance decision trees with AdaBoost and employ a BiLSTM model with Word2Vec and FastText embeddings for deeper semantic understanding. Our work compares these approaches to identify effective methods for emotion detection in Bangla, aiming to advance sentiment analysis in languages with fewer resources.**

***Key Words-* Emotion Analysis, Bangla Text, Machine Learning, Deep Learning, Sentiment Analysis.**

## I. Introduction

In the domain of computational linguistics, the analysis of emotions in text has been instrumental in addressing a wide array of challenges, particularly in English. Studies have successfully extracted emotions from suicide notes (Yang et al., 2012; Desmet and Hoste, 2013) [1], identified offensive sentences in dialogues (Allouch et al., 2018) [2], and offered support to cancer patients through sentiment analysis (Sosea and Caragea, 2020) [3]. The progress in these areas largely stems from extensive research dedicated to intricate, multi-label emotion detection tasks seen in initiatives like the SemEval Affective Texts (Strapparava and Mihalcea, 2007) [4], SemEval Effects of Tweets (Mohammad et al., 2018) [5], and GoEmotion (Demszky et al., 2020) [6]. These efforts have paved the way for significant advancements in understanding and interpreting human emotions through text.

Despite the progress in languages like English, Bangla, the sixth most spoken language worldwide[1] and the primary language in Bangladesh, has seen limited exploration in this domain. As Bangladesh emerges as a middle-income nation with expanding technology access even in rural areas (Basunia, 2022; Islam and Saeed, 2021) [7], the importance of analyzing Bangla text for emotional content becomes increasingly relevant. Such analyses can significantly impact social welfare and business sectors by providing insights into public sentiment and emotional responses.

This research paper explores the analysis of emotions within Bangla text, utilizing a dataset (Islam et al., 2022) [8] of 22,698 public comments from social media, covering diverse topics related to personal issues, politics, and health. We explore various machine learning models, such as Linear SVM, KNN, and Random Forest, applying n-gram and TF-IDF vectorization to capture the linguistic nuances across unigrams, bigrams, and trigrams.

Additionally, we investigate the role of PCA in dimensionality reduction, comparing results with and without its application. Our study extends to decision trees enhanced with AdaBoost, comparing their performance in emotion detection. Finally, we embrace deep learning through a BiLSTM model, enriched with Word2Vec and FastText embeddings, to gain a deeper understanding of semantic relationships in Bangla text. This comprehensive approach aims to establish a benchmark in emotion detection for Bangla, contributing to the broader field of sentiment analysis in under-researched languages.

## II. Literature Review

Rahman et al. [9] conducted a comprehensive investigation into discerning detailed emotions in Bangla text, transcending the binary categorization of sentiments into positive or negative. The authors meticulously curated and annotated comments from Bangla Facebook groups engaged in discussions on social and political topics. Emotions such as happiness, sadness, disgust, surprise, fear, and anger were analyzed. Employing machine learning techniques, including Support Vector Machines (SVM), the study achieved a commendable 53 percent accuracy in emotion classification. This research offers promising insights into the nuanced interpretation of emotions in Bangla text.

Yang et al. [1] addressed the challenge of unraveling emotions in written text, a pivotal task for sentiment analysis. The focus of their work was on participating in a competition dedicated to recognizing emotions in medical notes, particularly those associated with suicide. The proposed system amalgamated various language models, encompassing the identification of specific words, the detection of emotional cues within sentences, and the application of machine learning for emotion classification. By combining these diverse methods, the study aimed to achieve accurate detection and classification of 15 different emotions in suicide notes.

In a distinct domain, Allouch et al. [2] concentrated on developing an agent-based system tailored to assist children with Autism Spectrum Disorder (ASD) in navigating their social interactions. The research specifically addressed situations wherein ASD children unintentionally uttered insulting sentences. The authors compiled a dataset comprising insulting and non-insulting sentences with inputs from parents of ASD children. Machine learning methods, including SVM, Multi-Layer Neural Network, and Tree Bagger, were employed to predict insulting sentences with

[1]https://en.wikipedia.org/wiki/List$_of_languages_by_total_numbers$

precision and recall exceeding 75 percent. The study underscores the significance of automated agents in furnishing pertinent feedback and support for ASD children in social communication, especially in scenarios involving potential insults.

Carlo Strapparava et al. [4] employ a dataset consisting of 250 annotated headlines in the development set and 1,000 annotated headlines in the test set, sourced from major newspapers and news websites. The task revolves around the challenging classification of emotions and valence in these headlines, exploring the intricate connection between lexical semantics and emotions. The study features five distinct models, each contributing to the diverse landscape of automatic emotion recognition. These models include UPAR7, utilizing a rule-based linguistic approach; SICS, employing a word-space model and seed words; CLaC, presenting both a knowledge-based unsupervised approach and a supervised Naïve Bayes classifier (CLaC-NB); UA, utilizing statistics from web search engines; and SWAT, a supervised system employing a unigram model with synonym expansion. This ensemble of models reflects a comprehensive exploration of emotion recognition in affective text, offering insights into various methodologies and their performance in the given task.

Islam et al. [8] introduce a novel dataset for fine-grained emotion analysis in the Bangla language. The authors propose EmoNoBa, comprising 22,698 Bangla public comments from social media across 12 domains, labeled for six fine-grained emotion categories based on the Junto Emotion Wheel. They highlight the challenges of working with low-resourced Bangla language and the limitations of existing emotion datasets. The dataset is meticulously prepared to preserve linguistic richness and challenge classification models. The experiments involve baseline models, including linguistic features, recurrent neural networks, and pre-trained language models, with surprising results: random baselines outperform neural networks and pre-trained language models. The authors release the dataset and models for research purposes. The extensive data collection, annotation process, and detailed analysis of linguistic features provide valuable insights for future research in emotion detection for the Bangla language, emphasizing the need for improved contextual understanding in transformer-based models.

## III. Dataset

### A. Dataset Collection

The EmoNoBa dataset is a comprehensive collection of 22,698 public comments in Bangla, sourced from various social media platforms. This dataset was meticulously compiled to support research in emotion analysis within the Bangla-speaking community. It encompasses comments spanning 12 different domains, labeled for 6 different emotion categories **(love, joy, surprise, anger, sadness, fear)** including but not limited to personal, politics, and health, making it a rich and diverse resource for understanding emotional expression in digital communication. We conducted a stratified split that accounts for multiple labels per instance, dividing the data into a training set comprising 80% of the total, a testing set making up 15% of the total and the rest of the 5% goes into the validation set. This method ensured that each set contained a proportional representation of the various label combinations present in the overall dataset.

| **Categories** | **Data** |
|---|---|
| Love | কুমিল্লা বেষ্ট হবে |
| Joy | সত্যিকার মানুষ তারাই ভাই |
| Surprise | এটা কিভাবে কিনব |
| Anger | খেলতে না পারলে এটাই বলে বাংলাদেশ |
| Sadness | লকাল বাস ভালো এটা থেকে |
| Fear | যদি গড় গ্রেড সি চলে আসে |

Fig. 1. Sample Dataset

### B. Dataset Objective:

For six basic emotion categories, the aim is to figure out all the feelings expressed in a piece of text.

### C. Dataset Statistics and Analysis:

Our dataset consists of 22,698 entries. On average, each entry is about $1.36 \pm 0.82$ sentences long. The typical sentence length is roughly $11.70 \pm 10.70$ words. The majority of the data, 77.28% to be exact, comes from YouTube comments. Additionally, 15.3% of the entries express more than one emotion.

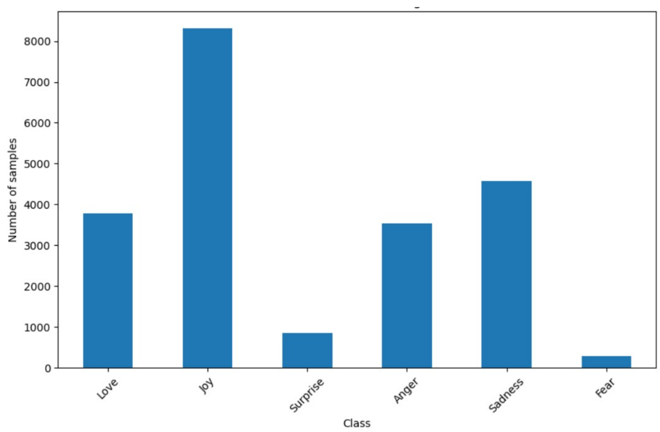


Fig. 2. Bar Chart of Dataset

Figure 2 presents a bar chart illustrating the distribution of samples across different emotional categories within the dataset. The graph depicts six emotions: Love, Joy, Surprise, Anger, Sadness, and Fear. From the visualization, we observe that 'Joy' has the highest number of instances, significantly more than the other emotions, followed by 'Sadness' which also has a large number of samples. 'Love' and 'Anger' are represented in a moderate number of instances. In contrast, 'Surprise' and 'Fear' are the least represented emotions in the dataset, with 'Fear' having the fewest instances. This distribution highlights the variability of emotional expressions in the dataset and indicates which emotions are more frequently expressed in the collected samples.

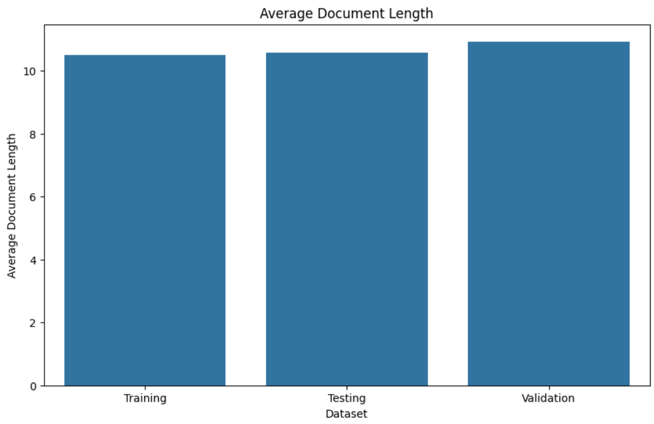


Fig. 3. Length of the Documents

Figure 3 illustrates the average length of the documents in each subset of the dataset, which appears to be consistent across the three subsets, indicating that the documents are, on average, of similar length whether they are part of the training, testing, or validation set.

### D. Word Cloud:


Fig. 4. Word Cloud

Figure 4 represents a word cloud which is a visual representation of the most frequently occurring words within the dataset used for our research. Each word's size is proportional to its frequency: the larger the word, the more often it appears in the dataset.

### E. t-SNE:

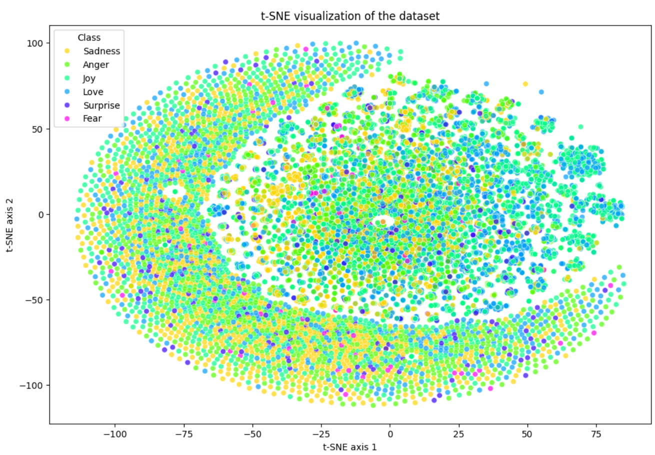


Fig. 5. t-SNE

Figure 5 shows a t-SNE visualization of our dataset, where we've transformed the text data into numerical values using TF-IDF (Term Frequency-Inverse Document Frequency) before applying t-SNE. The visualization clusters the data points based on their emotional content, with colors representing different emotions, revealing patterns and relationships between the various emotions expressed in the dataset.

## IV. Methodology

The methodology followed in our study is shown in Figure 6. Here are some of the following steps that we have implemented in our work.

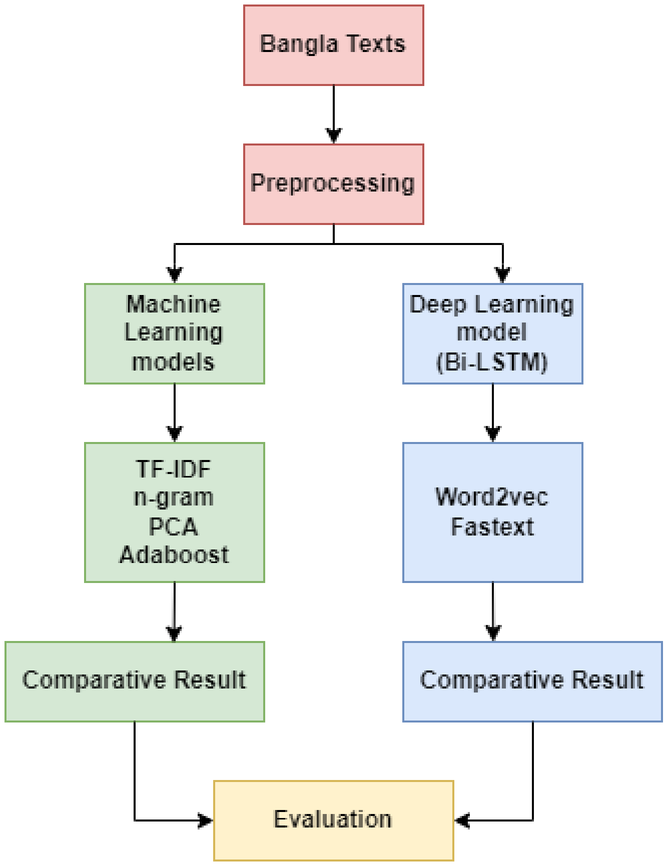


Fig. 6. Methodology

### Phase I: Dataset Collection

We collected our data from a dataset named as EmoNoBa, comprising 22,698 public comments in the Bangla language, meticulously curated from diverse social media platforms, with a predominant focus on YouTube.

### Phase II: Dataset Preprocessing

For preprocessing our dataset, we focused on cleaning the text to ensure consistency and accuracy for analysis. This involved removing all punctuation marks and emojis, as these elements could interfere with our emotion analysis algorithms. We also addressed the issue of multiple spaces within the comments, reducing them to single spaces to maintain uniformity in the text data. These steps helped to streamline the dataset, making it more straightforward for the computational models to process.

*Phase III: Classification*

*1) **Machine Learning Approach**:* Figure:7 shows an overview of our Machine learning approach. Our research adopted a comprehensive machine learning approach to classify emotions in text, leveraging three widely recognized models:**Linear Support Vector Machine (SVM), k-Nearest Neighbors (kNN),** and **Random Forest Classifier**. To refine our analysis and enhance the models' understanding of textual data, we utilized **Term Frequency-Inverse Document Frequency (TF-IDF)** in conjunction with n-grams—specifically unigram, bigram, and trigram configurations. TF-IDF is a statistical measure that evaluates how relevant a word is to a document in a collection of documents, emphasizing words that are frequent in a document but not across documents. This approach helps in distinguishing the significance of words in terms of their relevance to specific emotions being expressed. We applied

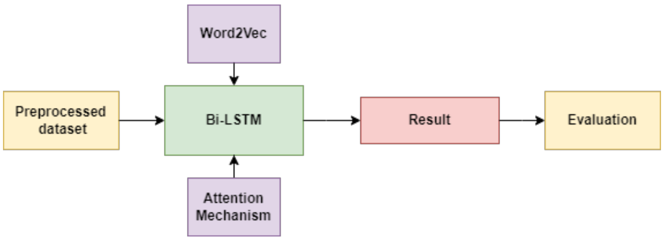


Fig. 7. Flow Chart of Machine Learning Approach

visual, textual, and multimodal approaches individually. The models are given below:

- **n-gram**: To assess the influence of text structure on model performance, we applied the aforementioned n-grams techniques, which consider single words, pairs, and triplets of consecutive words as unique features. This method allows for a nuanced analysis of the linguistic patterns that may denote emotional states.
- **PCA**: With an aim to explore the potential benefits of dimensionality reduction on our models' predictive accuracy, we implemented Principal Component Analysis (PCA). PCA simplifies the complexity of high-dimensional data while preserving as much variability as possible, facilitating a more streamlined and potentially more insightful analysis.
  Our experimentation yielded noteworthy outcomes. For the Linear SVM model, integrating TF-IDF with a unigram approach—without applying PCA—resulted in the highest macro-average F1 score of 0.63, marking the most effective combination for this model in identifying the six emotions under study. Similarly, the kNN model achieved its highest micro-average F1 score, a measure of overall accuracy, at 0.57 for the TF-IDF enhanced unigram configuration without PCA. This setup proved most proficient in emotion classification across all tests conducted with the kNN model. The Random Forest model mirrored these results, attaining its optimal performance with a TF-IDF supported unigram setup without PCA, achieving a macro-average F1 score of 0.57, which was the highest among the tested configurations for this model.
- **AdaBoost**:Furthermore, we extended our work to include an experiment with AdaBoost in conjunction with a Decision Tree model to evaluate its impact on classification performance. AdaBoost, a boosting algorithm, aims to improve the accuracy of a given model by combining multiple weak learners into a stronger learner. Before the implementation of AdaBoost, the Decision Tree model achieved an F1 score of 77.99. After applying AdaBoost, there was a slight improvement in performance, with the F1 score increasing to 78.6. This increment underscores the effectiveness of AdaBoost in enhancing the model's predictive accuracy. Figure-7 shows an overview of our approach.

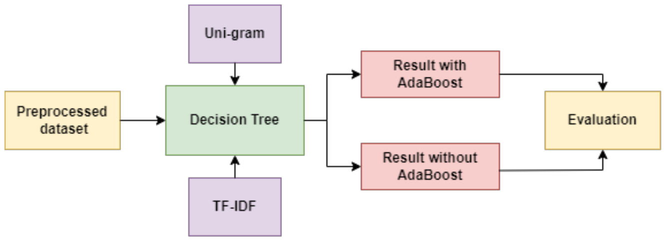


Fig. 8. Flow Chart of AdaBoost Analysis

*2) **Deep Learning Approach**:* Figure-9 shows an overview of our Deep Learning approach. For our deep learning approach, we opted for a Bidirectional Long Short-Term Memory (BiLSTM) network enhanced with Word2Vec embeddings (Mojumder et al.) [12]. The BiLSTM model, renowned for its proficiency in capturing sequential information from text in both forward and backward directions (Hochreiter et al.) [10], served as the backbone of our architecture. In addition to the BiLSTM layer, we integrated an attention mechanism (Bahdanau et al.) [11] to refine our model's focus. Our implementation included the initialization of word vectors using Word2Vec, which leverages the context of words in a corpus to produce word embeddings. Combining the strengths of Bi-LSTM, Word2Vec, and attention mechanisms, we aimed to create a model capable of effectively discerning and classifying the emotional undertones of text data.

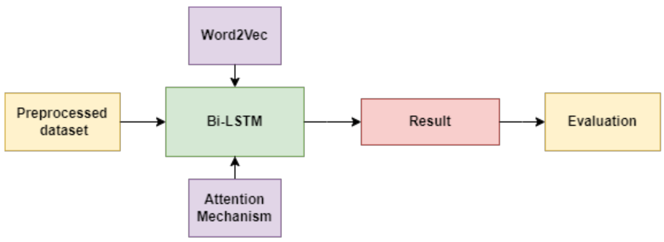


Fig. 9. Flow Chart of Deep Learning Approach

*Phase IV. Evaluation*

To evaluate the performance of our classification model, we deployed a comprehensive set of metrics, including accuracy, precision, recall, and the F1 score. These metrics were calculated in three distinct forms to provide a nuanced view of the model's performance: macro, micro, and weighted averages. Furthermore, we calculated the **macro-average F1 score** across all emotions. Table-I Shows our evaluation metrics.

TABLE I
EVALUATION METRICS

| Metrics | Micro | Macro | Weighted |
|---|---|---|---|
| Accuracy | N/A | N/A | N/A |
| Precision | Yes | Yes | Yes |
| Recall | Yes | Yes | Yes |
| F1-score | Yes | Yes | Yes |

## V. EXPERIMENTAL RESULTS

### A. *Machine Learning Based Results*

TABLE II
LINEAR SVM BASED RESULTS

| n-Gram | F1-score |
|---|---|
| Uni-gram (PCA) | 0.56 |
| **Uni-gram** | **0.63** |
| Bi-gram (PCA) | 0.5 |
| Bi-gram | 0.57 |
| Tri-gram (PCA) | 0.46 |
| Tri-gram | 0.49 |

Table-II presents the F1 scores for different n-gram models using a Linear SVM classifier. It shows that the unigram model without PCA achieved the highest score of 0.63, indicating it was the most effective at classifying emotions in our dataset. Other models, including those with PCA, showed lower effectiveness with F1 scores ranging from 0.49 to 0.57.

TABLE III
KNN BASED RESULTS

| n-Gram | F1-score |
|---|---|
| Uni-gram (PCA) | 0.54 |
| **Uni-gram** | **0.57** |
| Bi-gram (PCA) | 0.52 |
| Bi-gram | 0.55 |
| Tri-gram (PCA) | 0.46 |
| Tri-gram | 0.46 |

In Table-III, we see the performance of various n-gram models using the kNN classifier. Similar to the Linear SVM results, the unigram model without PCA performed the best with an F1 score of 0.57. The application of PCA seems to have slightly reduced the effectiveness of the unigram and bigram models, with trigram models consistently scoring lower, both with and without PCA.

TABLE IV
RANDOM FOREST BASED RESULTS

| n-Gram | F1-score |
|---|---|
| Uni-gram (PCA) | 0.55 |
| **Uni-gram** | **0.57** |
| Bi-gram (PCA) | 0.52 |
| Bi-gram | 0.52 |
| Tri-gram (PCA) | 0.45 |
| Tri-gram | 0.44 |

This Table-IV outlines the F1 scores for the Random Forest classifier across different n-gram configurations. The highest F1 score was seen in the unigram model without PCA at 0.57, suggesting that simpler models may perform better with this dataset and classifier. The addition of PCA did not significantly improve the F1 scores across unigram, bigram, and trigram models.

TABLE V
DECISION TREE BASED RESULTS

| AdaBoost | Macro average F1-score |
|---|---|
| With AdaBoost | 78.6 |
| without AdaBoost | 77.99 |

Table-V shows the macro average F1-scores for a Decision Tree model with and without the use of AdaBoost, which is a technique to enhance classification accuracy. With AdaBoost, the model achieved a higher F1-score of 78.6, compared to 77.99 without AdaBoost. This suggests that AdaBoost provided a slight improvement in the model's ability to classify emotions.

### B. *Deep Learning Based Results*

TABLE VI
DECISION TREE BASED RESULTS

| Moel | Macro average F1-score |
|---|---|
| Bi-LSTM | 23.73 |

Here, Table-VI shows the results for a Bi-LSTM deep learning model are presented. This model utilized Word2Vec for word embeddings and achieved a macro average F1-score of 23.73. The score reflects the average effectiveness of the model across different emotional categories within the dataset. Despite being a sophisticated deep learning approach, the F1-score indicates there is significant room for improvement in the model's performance.

## VI. RESULT ANALYSIS

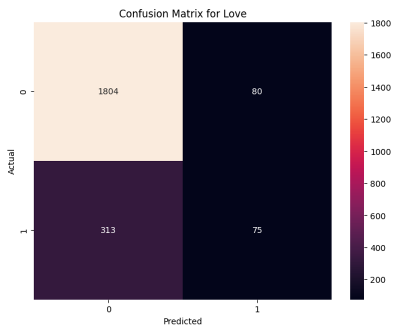


Fig. 10. Confusion Matrix of Love

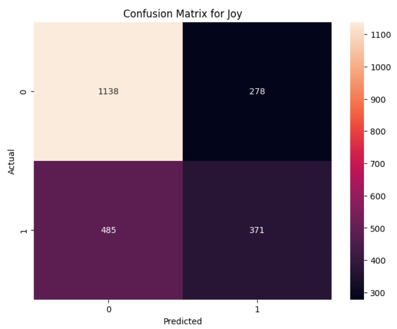


Fig. 11. Confusion Matrix of Surprise

Figure-10,11 shows us a sample of confusion matrix of 2 labels for our best performing model (Decision Tree).

TABLE VII
RESULT ANALYSIS OF EMOTION CLASSIFICATION MODELS

| Model | n-Gram | PCA | AdaBoost | F1-score |
|---|---|---|---|---|
| SVM | Uni-gram | No | No | 0.63 |
| KNN | Uni-gram | No | No | 0.57 |
| Random Forest | Uni-gram | No | No | 0.57 |
| **Decision Tree** | Uni-gram | No | Yes | **78.6** |
| Bi-LSTM | No | No | No | 23.73 |

Table-VII provides a summary of the F1-scores for various emotion classification models. The Linear SVM model, using a unigram approach without PCA or AdaBoost, scored the highest with 0.63. The kNN and Random Forest models also used unigrams without PCA or AdaBoost and scored slightly lower, with F1-scores of 0.57. The Decision Tree model showed a significant improvement when AdaBoost was applied, jumping to an F1-score of 78.6 compared to models without AdaBoost. The Bi-LSTM model, which did not use PCA or AdaBoost, had an F1-score of 23.73, indicating that while it might capture the nuances of language with Word2Vec, it was less effective at emotion classification compared to the other models.

## VII. LIMITATIONS AND FUTURE WORKS

Our study, while insightful, had some limitations. The data came only from social media, which doesn't fully capture how people use Bangla in everyday life or in different contexts. Also, some emotions were more common in the data than others, which could have affected our results. Plus, the tools we used to analyze the text, like n-grams, might not have caught all the nuances of the language. In future research, we aim to use a wider variety of texts to get a fuller picture of how Bangla is used. We also want to balance out the data so all emotions are fairly represented. We're looking to try out newer technologies like BERT or GPT, which might be better at understanding context. Improving how we process the text before analysis could also help. Exploring other methods to combine different models might give us deeper insights into Bangla and emotions.

## VIII. CONCLUSION

In conclusion, our research took important steps in understanding emotions in Bangla text through social media comments. We used several machine learning models and found that the linear SVM model with unigram analysis worked best without reducing the data's complexity. Our study adds to the field of sentiment analysis in Bangla and points to promising directions for future research to improve accuracy and understanding of emotional expressions in language.

## REFERENCES


[1] Hui Yang, Alistair Willis, Anne De Roeck, and Bashar Nuseibeh. 2012. A hybrid model for automatic emotion recognition in suicide notes. Biomedical informatics insights, 5:BII–S8948. Journal, Year.

[2] Merav Allouch, Amos Azaria, Rina Azoulay, Ester Ben-Izchak, Moti Zwilling, and Ditza A Zachor. 2018. Automatic detection of insulting sentences in conversation. In 2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE), pages 1–4. IEEE.. Journal, Year.

[3] Tiberiu Sosea and Cornelia Caragea. 2020. Canceremo: A dataset for fine-grained emotion detection. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8892–8904.

[4] Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 70–74.

[5] Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018a. Semeval-2018 task 1: Affect in tweets. In Proceedings of the 12th international workshop on semantic evaluation, pages 1–17.

[6] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. arXiv preprint arXiv:2005.00547.

[7] Sazzad Reza Basunia. 2022. E-commerce in rural bangladesh: The missing dots. The Business Standard.

[8] Islam, Khondoker Yuvraz, Tanvir Hassan, Enamul Islam, Md Saiful. (2022). EmoNoBa: A Dataset for Analyzing Fine-Grained Emotions on Noisy Bangla Texts.

[9] Md Rahman, Md Seddiqui, et al. 2019. Comparison of classical machine learning approaches on bangla textual emotion analysis. arXiv preprint arXiv:1907.07826

[10] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation, 9(8):1735–1780.

[11] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

[12] Mojumder, Pritom Hasan, Mahmudul Hossain, Faruque Hasan, K. M.. (2020). A Study of fastText Word Embedding Effects in Document Classification in Bangla Language. 10.1007/978-3-030-52856-0$_35$.