Vulgar Speech Detection in Bangla Spoken Language

By

Hasibul Hasan	**ID: 011181117**
Tahsin Habib Brinto	**ID: 011181107**
Rokibul Islam	**ID: 011181102**
Chion Ghosh	**ID: 011181110**
Shahzadi Khadija Khan Himu	**ID: 011181105**

Submitted in partial fulfilment of the requirements
of the degree of Bachelor of Science in Computer Science and Engineering

February 28, 2022

Department of Computer Science and Engineering
United International University

Abstract

In recent times, the prevalence of abusive and vulgar language on online platforms seems to have become a cause for worry. However, in languages like Bengali, research on the frequency and identifying of vulgar language has largely remained neglected. In this paper, we provide analysis and an approach for detecting vulgarity in a Bengali sentence. Finding enough Bengali vulgar words to build a rich dictionary is difficult and because of less research on this topic, almost no data is readily available. Our data corpus comprises 19908 instances of Bengali plaintext that we have collected and created. We have experimented with various deep leaning approaches, such as Recurrent Neural Network (RNN). Our research finds that bidirectional RNN based two layers of Long Short-Term Memory (BiLSTM) with dropout layers to counter over-fitting provide the best results, considering both accuracy and loss of the validation data. This paper will help later researches on Bengali vulgar/profanity detection and provide a better, safer environment for counteracting targeted hate, bullying, and vulgar use of Bengali language.

Acknowledgements

This work would have not been possible without the input and support of many people over the last two trimesters. We would like to express our gratitude to everyone who contributed to it in some way or other. First, we would like to thank our academic advisors, Professor Dr. Mohammad Nurul Huda(PhD), United International University. Our sincere gratitude goes to Professor Dr. Md. Motaharul Islam(PhD) Dept. of CSE, United International University. We are also thankful to our every team member. Last but not the least, We owe to our family including our parents for their unconditional love and immense emotional support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Speech is the most ancient form of communication, and this is the most appropriate form of expressing human emotions. Though the beauty of speech is limitless, when it comes to hate speech, speech works as a weapon or acts as a slow poison to the victim or to the society.
Hate speech is defined as "public communication that expresses hatred or advocates violence toward a person or a group based on anything such as race, religion, or sexual orientation," according to the Cambridge Dictionary and Wikipedia [1]. Vulgar speech or vulgarity in a sense of speech is considered when the speech is in a more offensive and abusive state.
In order to detect and prevent vulgarity from speech through a machine, Automatic Speech Recognition (ASR) technology works as a torchbearer. Speech recognition technology is simply converting human spoken language into machine readable format, recognizing every part of that speech (example: words or phrases) and finally written out into the human readable text. By using machine learning algorithms this technology has gained the ability to recognize words and phrases from a speech with higher accuracy. This technology is under the field Natural language processing (NLP), a branch of Computer Science.

1.1 Problem Definition

In this era of fast growing technology, everyone is somehow connected with people on the internet. People are spending most of their time online for different purposes. There is a report which has stated that in 2018, the number of internet user was more than 4 billion with an increasing rate 7 percent. Social media user on that year was 3.2 billion and increasing 13% each year, and the mobile phone users were 5.1 billion, where the total population of the world was 7.593 billion on that year[2]. Definitely, it has a huge impact on the mankind. In Bengalidesh, here people are also accepting online platforms rapidly. Without any question this is definitely a good sign, but this blessing of technology can

turn into a social threat within a fraction of time, when the people use the technology in a wrong manner.
People of different ages around the world are facing online bullying, attacked by hate speech or vulgar speech in any platform. Cyberbullying in the internet or online verbal-harassment, spreading and frequently use of vulgar speech speech has begun to globalize and become a social problem. This ill-behaviour of some immature people causes tremendous psychological harm to the people of the society, beyond imagination. It's creating a threat to the healthy development and mental growth of human beings. It's gradually destroying people's mentality.
Based on this problem, In our research we're going to solve this real life social problem as well as this paper is aimed to provide a solution to detect Bengali vulgar speech with some Deep Learning algorithm and some supervised machine learning models.

1.2 Motivation

In Bengalidesh, we're frequently being harassed by vulgar speech, bullied and attacked by hate speech irrespective of our ages. In addition, this is one of the major causes for the fast growing suicide rate in this country. A case-study stated that nearly 49% Bengalideshi students have reported being a victim of cyberbullying. That report also cited that majority of those victims are female and this situation is getting worse day-by-day as outstanding development has been occurring in the communication technology sector in Bengalidesh. [3].
Bengali is one of the major spoken languages, and ranked 7 th in the world by its large number of speakers [4]. So Bengali is certainly becoming a more popular medium of communication over this vast internet.
Since it is an international issue, a substantial amount of research has been conducted to develop solutions for identifying vulgar speech in international language English, ranging from simple machine learning models to complicated deep neural network models. But a very few amount of research has been done and published to detect and prevent vulgar speech in Bengali language.
Since the huge number of user communicating with others through this enormous internet technology, we believe that, If a automated system can detect and prevent the vulgar words from a delivered speech and stop people from spreading any hate speech, we can make our society a better place with a full of healthy minded people.
Based on this motivation we are doing research on how we can make our society free of vulgar speech, help to reduce hate speech and give our future generation a polite environment so that they can grow their beautiful minds and learn the good things despite learning bad language.

1.3 Objectives

The primary objectives and the project outcomes of this research are:

- To design a model to recognize Bengali speech with more accuracy.
- To achieve high accuracy to detect vulgar or hate words as well as vulgar speech in real time.
- To improve the performance of our algorithm than the existing Bengali speech recognition.
- To create a scope for further research, study and progress in this specific area.

1.4 Contribution

The lack of a comprehensive user-friendly toolbox and a substantial voice corpus are two main challenges in doing research in Bengali Automatic Speech Recognition (ASR). Developing a user-friendly toolbox as well as a standard-sized Bengali speech corpus is a huge task for us.
The Recurrent Neural Network (RNN), among other methods, is widely popular and is utilized in machine learning applications where sequence modeling is critical. Other approaches have various drawbacks, such as lesser performance or a computationally difficult algorithm. As a result, combining the RNN approach with Long Short-Term Memory (LSTM) to acquire articulatory information appears to be a significant step forward in Bengali speech research.
We'll create a scope for further research on this specific problem. We believe that in the near future, more researchers will do research on Bengali language processing. The models and algorithms can be more optimized through more research.

1.5 Methodology

We are going to use Long Short-Term Memory(LSTM) for sentiment classifying, processing and making predictions for data. We will use Recurrent Neural Network(RNN) as learning models, classification analysis. As our problem is a classification problem and the task of our model is to detect whether the speech is vulgar or non-vulgar based on the supervised labeled dataset.

1.6 Organization of the Report

This paper is segmented into different chapters and different sections.

- Chapter-2: This chapter will contain the preliminaries and the literature review and gap analysis.
- Chapter-3: The Topic of this chapter will be requirement analysis, methodology design and the project plan.
- Chapter-4: The implementation and the results will be demonstrated in this chapter.
- Chapter-5: This chapter will contain all the standards, complex engineering problem mapping and the design constraints.
- Chapter-6: Finally this chapter will conclude and summarize this paper.
- References: All the references we used for writing this paper is stated in this section.

Chapter 2

Background

Speech recognition is a multidisciplinary field in computer science and computational linguistics that develops approaches and technologies that allow computers to recognize and translate spoken language into text. Automatic speech recognition (ASR), computer voice recognition, and speech to text (STT) are some of the terms used to describe it.

2.1 Preliminaries

Reader of this paper should have knowledge about algorithms, compiler design principles, computer science, computational linguistics.

2.2 Literature Review

For the past few years many researchers have done appreciable work on this amazing field, Natural Language Processing(NLP) and contributed to the research community with a number of great research containing amazing models and algorithms.
Anna Schmidt and Michael Wiegand They collaborated on a survey study that gives a concise, thorough, and systematic review of automatic hate speech identification, as well as a systematic analysis of existing techniques, with a focus on feature extraction in specific. It is mostly intended for NLP researchers who are novice to the topic of hate speech detection who wish to learn about current research. We learned from that paper that character-level approaches outperform token-level ones. Various complicated features that need additional linguistic expertise, such as relationship parsed information, or features that mimic specific linguistic entities, such as imperatives or etiquette, have also been demonstrated to be effective. [5].
Another works on detecting abusive Bengali text has done by Estiak Ahmed Emon, Shihab Rahman and their team, which is basically a Deep Learning approach. We observed that they tokenized their data from a dataset. From where they processed the data into root form, then they used that root data to train their algorithm. For tokenization they used count-vectorizer, tf-idf vectorizer, word embedding. Which gives statistically

82.2% accuracy [6]. The study goal of that paper was to test Logistic Regression (Logit), Linear Support Vector Classifier (LinearSVC), Random Forest (RF), Multinomial Nave Bayes (MNB), and Recurrent Neural Network (RNN) with a Long Short Term Memory (LSTM) cell to detect multi-type abusive Bengali text. In addition, they've introduced new stemming methods for Bengali, which aid in improving algorithm efficiency.
Alex Graves, Abdul Rahman Mohamed and Geuffrey Hinton have done a research on speech recognition. This paper introduced the use of Long Short-Term Memory(LSTM) layers, RNNs and assess it's emergence for speech recognition. RNNs may also benefit from spatial depth. They employed a recently developed end-to-end learning strategy in which two distinct RNNs are concurrently trained as acoustical and lexical models. On the Acoustic-Phonetic Continuous Speech Corpus/TIMIT, a combination, bidirectional Long Short-term Memory cell, end-to-end training, and weight noise produces excellent performance in phoneme recognition. [7].

2.2.1 Similar Applications

We explored and found some amazing application similar to our interest, which are:

- Google Assistant
- Siri - Apple
- DeepSpeech - Mozilla
- Wave2Letter - Facebook
- OpenSec2Sec - NVIDIA
- TensorFlow ASR

2.3 Gap Analysis

Here we summarised the gaps where we intend to work and improve the performance.

In such a case, data cleaning will be the most important factor in achieving good performance. We'll concentrate on cleaning the data while keeping the Bengali language structure in mind, and we'll employ specialist tools and modules for Bengali text data cleaning and preprocessing. To counteract the over-fitting issue, we can add dropout layers, reduce the network capacity and use weight regularization.

2.4 Summary

We have observed, analyse, experiment other's work and get the knowledge about how they've done it, and we've come to an abstraction about how can we make our research done properly.

Chapter 3

Project Design

3.1 Requirement Analysis

Requirements analysis, commonly known as requirements engineering, is a method for determining a new product's needs and expectations. It entails constant communication with the product's stakeholders and end-users to establish expectations, manage issues, and document all of the product's important requirements.

3.1.1 Functional and Nonfunctional Requirements

Functional Requirement: What should this system must do is its functional requirement.

- This system will detect the vulgar words from an input speech

Non-Functional Requirements:

- Convert an input speech to a text.
- Breakdown the whole text into words.
- Classify the text using our model and outputs the result.

3.2 Detailed Methodology and Design

For our model, we will use two layers of bidirectional Long Short Term Memory cell (LSTM) with a dropout layer with 0.5 as the drop rate in-between. This will reduce the network capacity and prevent the model from over-fitting issue. The two more dense layers will be added and finally the output layer will be a single node dense layer with sigmoid activation method.

We also considered developing two other models with single layer of LSTM and a Gatted Recurrent Unit (GRU) layer and 1D Convolutional Neural Network. Based on our experiments LSTM has shown remarkable results and metrics among all of these, specially for text of sequence classification problems.

3.3 Project Plan

Project plan for FYDP-1 and FYDP-2 are shown below:

Figure 3.1: FYDP-1 Timeline Gantt Chart

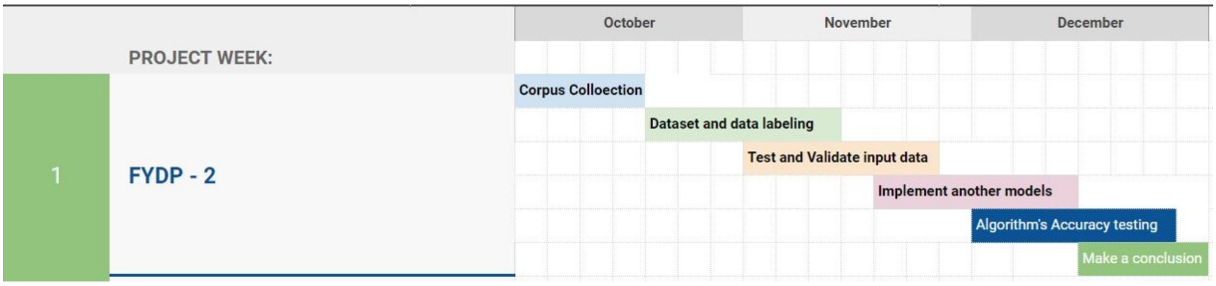


Figure 3.2: FYDP-2 Timeline Gantt Chart

3.4 Summary

After reviewing a lot of related research papers and books, we have selected our step by step methodology and we are going to develop our model with two Bidirectional LSTM layers and will take proceedings to counter over fitting issue as our dataset is relatively smaller and repetitive.

Chapter 4

Implementation and Results

In this chapter we discuss how we collect our dataset, pre-process and clean those data to train our model. We also demonstrate our methodology for vulgarity detection in speech, we provide detailed experiments and show the result comparison among different models. We experiment with different models to see the different performance and results which are provided in this chapter.

4.1 Data set

Collecing and finding our desired dataset containing specially Bengali vulgar words have been the most challenging part for us. We've collected the dataset contains Bengali vulgar words that are racist, slang, sexiest and offensive.[8] These Bengali speech are collected from sources like wikipedia, news articles, newspapers, facebook comments, and many other online platforms [9] [10] and 1,000 data manually collected by us from several places. Our dataset contains more than 20,000 raw text data which are labeled as vulgar and non-vulgar. The label is classified as 0 which is non-vulgar and 1 which is vulgar.

4.2 Data Preprocessing

As we have collected our data from different places, and a portion of the dataset is created by us. So, it's important to preprocess the data to build a model. In this section we used nltk and bnlp(nlp toolkit for bengali words) modules to preprocess and clean the data. We have used several preprocess technique to clean the dataset. We use stemming algorithm to reduce the similar words. We apply this technique to find the root word. Stemmer is very important in nlp to tokenize the words from input stemmer. After that we have used lemminzation to grouping together the warped forms of word to analysed as single item. So in this way we links similar meanings of word with another word. For Bangla stopwords, punctuation removed, we used bnlp. Through the bnlp methods we removed the stopwords, punctuation, emoji , hyperlink from our bengali words to tokenize the bengali words efficiently.

4.3 Implementation

Our aim is to classify a speech into a binary class classification wheather it contains any vulgarity or not, that's why using some pre-processing techniques we clean the data, removing unnecessary characters, punctuations, emojis etc and after doing that we've done some language processing tasks.

Tokenization and Padding Sequence : Tokenization is the process where the plain text converts into a token value and that uses a database to store the mapping. After preprocessing the data we tokenized the word from the sentences.Tokenization is a very important part when we handle the text data. As we work with text data first we have splitted words from a sentence. We have splitted sentences into smaller units, such as individual terms. We have used the tensorflow module to tokenize the words. This class allows to vectorize the text data and turning the sentences or phrases into a vector where the coefficient for each token could be binary, based on word count, based on tf-idf.

Padding sequence When all of the inputs are the same length, this is known as a padding sequence. We have padded our tokenized words.we have used pad sequence to ensure that all words in a vector have the same length.we define our pad length 120.

Word Embedding : Word embeddings are a sort of word representation that allows for the depiction of words with comparable meaning. We have embedded our words, in this way each word is represented by a real-valued vector with many dimensions, often tens or hundreds.

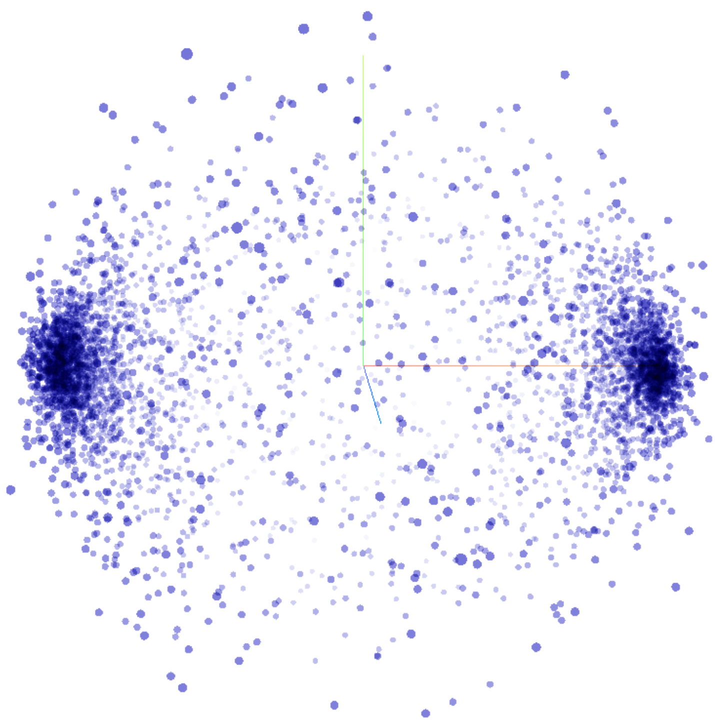


Figure 4.1: Embedding projection

Recurrent Neural Network (RNN) : After prepossessing and embedding we apply the Rnn(Recurrent neural network) algorithm in our prepossessed dataset. As we know, RNNs are useful for problems where the sequence of events is more essential than the individual elements.RNNs contain a kind of internal memory that allows earlier inputs to influence future predictions. If we know what the preceding words were, it's much easier to predict the following word in a phrase with greater accuracy.

Long Short-Term Memory (LSTM) : We have used bi-directional long short time memory (BiLSTM). The LSTM approach relies on Recurrent Neural Networks (RNN) and is a deep learning architecture. It remembers the previous state, and merges it with the next state. A multiple word string can be used in LSTM to determine which class it belongs to. When working with natural language processing, this is quite beneficial. The model will be able to discern the true meaning of the input text and offer the most correct output class if enough degrees of word embedding, and encoding are applied in LSTM.

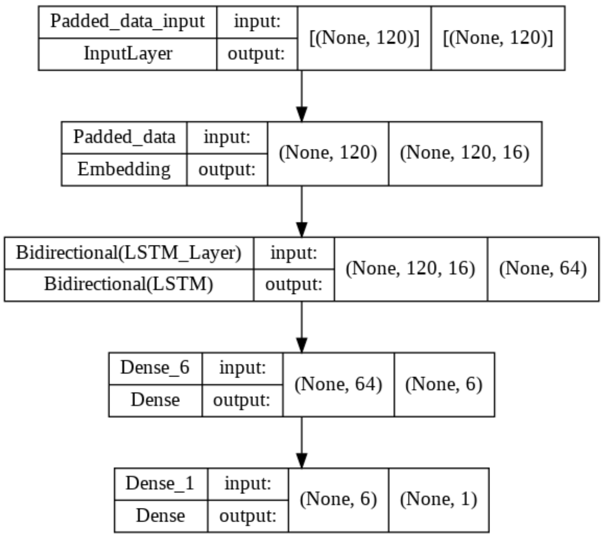


Figure 4.2: Single Layer LSTM

We used a dropout layer with drop-rate of 0.5 to regulate the only LSTM layer, and keep it as free of bias as feasible.

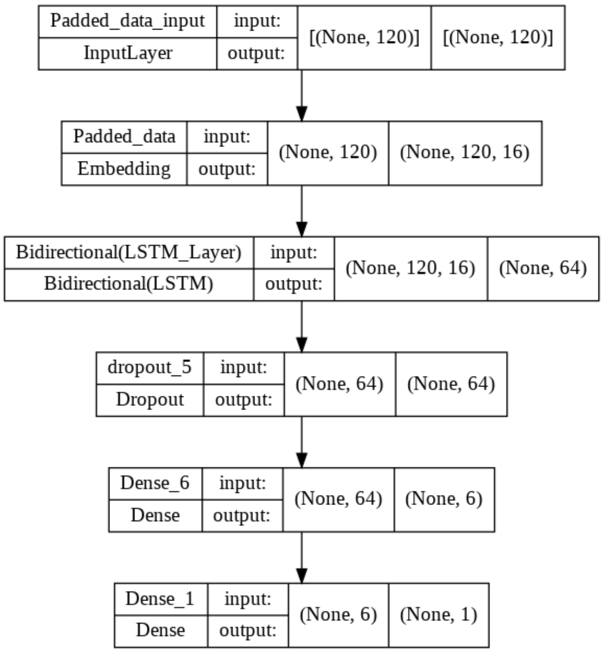
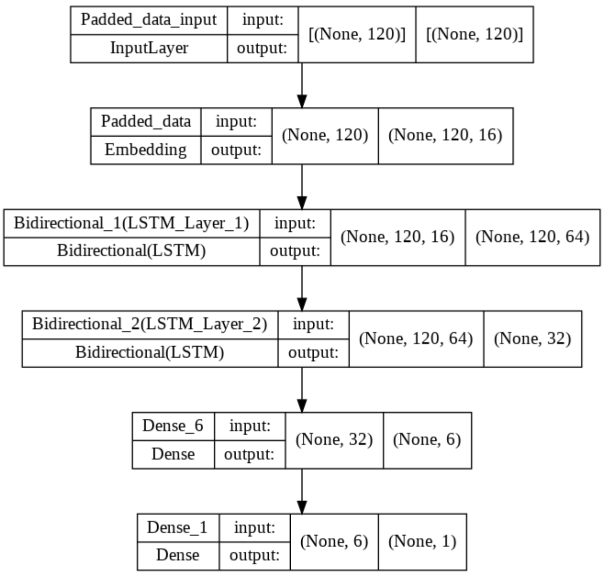
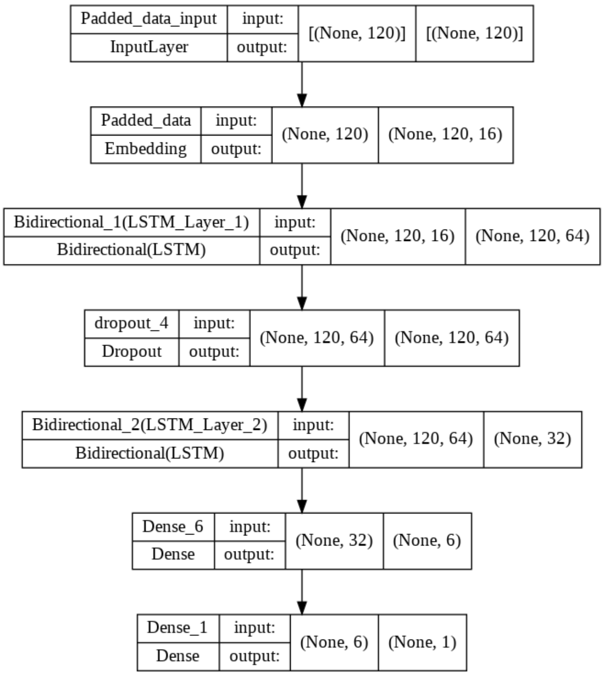


Figure 4.3: Single Layer LSTM (Reduced)



Again we have used a dropout layer with a drop-rate of 0.5 to regulate our double layer LSTM network.



4.4 Results and Discussion

We're demonstrating the results and performance evaluation we've observed by using several models. We consider the loss function and accuracy measure as metric for our model. we work on our validation set to get these metric.

4.4.1 Single Layer LSTM Model:

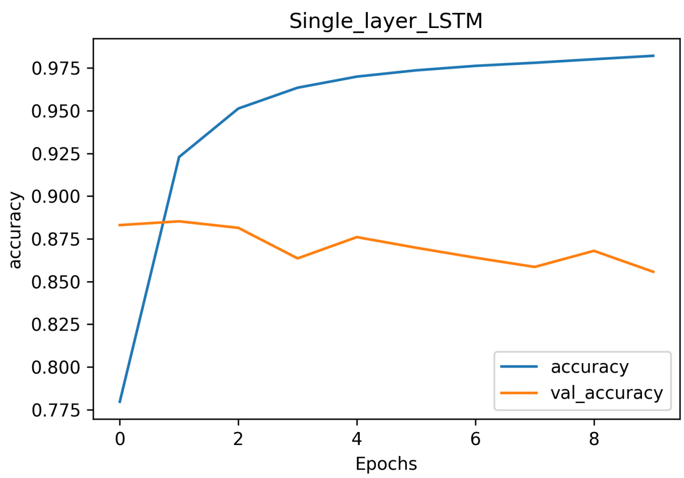


Figure 4.6: Single Layer LSTM accuracy

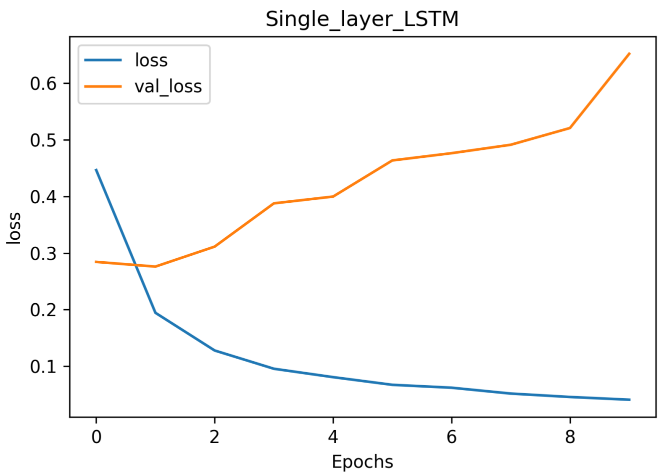


Figure 4.7: Single Layer LSTM loss

4.4.2 Single Layer LSTM Model (Reduced):

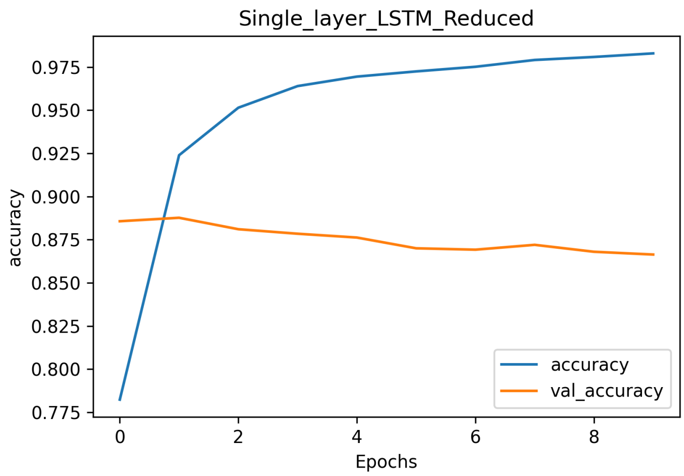


Figure 4.8: Single Layer LSTM accuracy(Reduced)

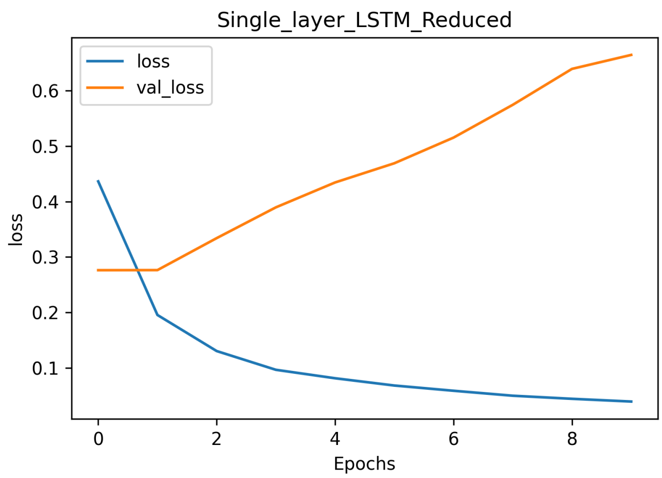


Figure 4.9: Single Layer LSTM loss(Reduced)

4.4.3 Double Layer LSTM Model :

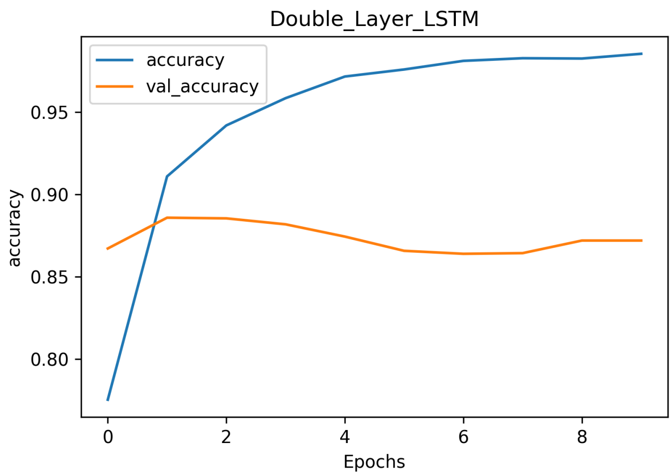


Figure 4.10: Double Layer LSTM accuracy

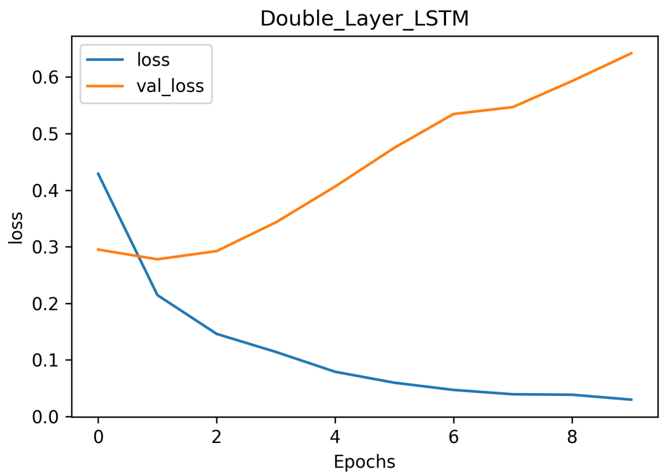


Figure 4.11: Double Layer LSTM loss

4.4.4 Double Layer LSTM Model (Reduced) :

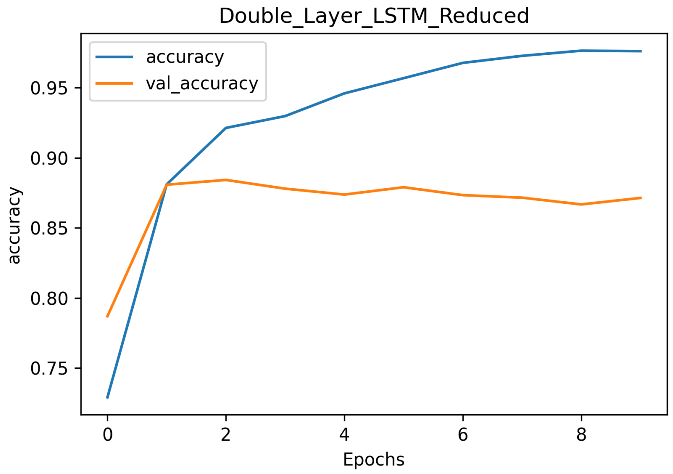


Figure 4.12: Double Layer LSTM accuracy (Reduced)

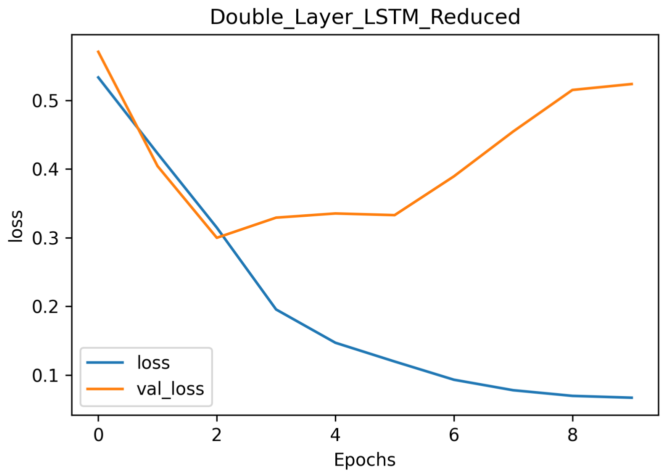


Figure 4.13: Double Layer LSTM loss (Reduced)

4.5 Summary

We observed that Reduced Double Layer LSTM model is performing better among the other models, it gives the 87% accuracy and the loss is 0.48. If we improvise our dataset, then the results will be better than the current version.

Chapter 5

Standards and Design Constraints

A standard is a method of doing anything that is repeatable, harmonized, agreed upon, and documented. Detailed specifications or other precise criteria are contained in standards, which are intended to be applied systematically as a rule, direction, or definition. A design constraint is a restriction on the objectives or operating conditions that a system is anticipated to perform under. A design constraint can have an impact on the system's operation characteristics and functioning, for example.[11].

5.1 Compliance with the Standards

5.1.1 Software Standards and Technologies

We're using,

- Speech: (dot).wav format
- Text: unicode(utf-8)
- Programming Language: python
- Library/framework: sklearn, numpy, pandas, TensorFlow
- Data importing and exporting: csv formate.
- Model: Recurrent Neural Network(RNN), Long Short Term Memory(LSTM)
- Environment: python 3

5.2 Design Constraints

5.3 Complex Engineering Problem

Complex engineering problems, as defined by the Washington Accord (IEA2015) [12], are those that:

- Without in-depth engineering understanding, this problem cannot be solved.
- Involve a variety of technical, engineering, and other issues that are broad or contradictory.
- There is no obvious solution, and developing appropriate models necessitates abstract thinking and inventiveness in analysis.
- Involve issues that aren't commonly encountered.
- Professional engineering standards and codes of practice cover a wide range of issues.
- Involve a variety of stakeholders with a wide range of needs.
- Many component pieces or sub-issues are included in high-level problems.

5.3.1 Solving Complex Engineering Problems

P1 and some or all of P2 to P7 are characteristics of Complex Engineering Problems..

- Cannot be resolved without in-depth engineering knowledge at the level of one or more of k3, k4, k5, k6, k8, which allows a fundamental-based, first principle analytical approach. Our research necessitates a deep understanding of the knowledge profile.
- Depth of analysis is necessary (P3): There is no clear solution, thus abstract thinking and inventiveness in analysis are required to create appropriate models.
  - Our research necessitates abstract thinking, which is defined as the ability to absorb information from our senses and connect it to the rest of the world. [13].
  - In order to construct appropriate models, our study preserves originality in analysis.
- Familiarity of Issues(P4): Many other languages, such as English, Arabic, and others, have done a lot of good work on vulgar speech recognition. However, there hasn't been as much progress in the Bengali language. As a result, our research focuses on a rarely encountered problem.
- Extent of applicable codes(P5): Outside concerns are covered by professional engineering standards and rules of practice, and our research upholds these standards and codes of engineering ethics.
- Extent of stakeholder involvement and conflicting requirements(P6): Our study includes a varied set of stakeholders with a wide range of demands.
- Interdependence(P7): A variety of subsystems are involved in our research.

Table 5.1: Mapping with complex engineering problem solving.

P1 Depth of Knowledge	P2 Range of Conflicting Requirements	P3 Depth of Analysis	P4 Familiarity of Issues	P5 Extent of Applicable Codes	P6 Extent of Stakeholder Involvement	P7 Interdependence
√		√	√	√	√	√

Table 5.2: Mapping with knowledge profile of P1 in complex problem solving.

K	Profile-Name	Remark
K1	Natural Science	
K2	Mathematics	
K3	Engineering Fundamentals	√
K4	Specialist Knowledge	√
K5	Engineering Design	√
K6	Engineering Practice	√
K7	Comprehension	
K8	Research Literature	√

5.3.2 Complex Engineering Activities

Complex activities are (engineering) activities or projects that include some or all of the activity domain's features (A1 - A5)

- A1-Range of resources: Make advantage of a variety of resources.
- A3-Innovation: Incorporate unique applications of engineering principles and research-based knowledge.
- A4- Consequences for society and the environment: Have major implications in a variety of situations, and are difficult to forecast and mitigate.
- A5- Familiarity: Apply principles-based ways to go beyond existing experience.

Table 5.3: Mapping with complex engineering activities.

A1 Range of resources	A2 Level of Interaction	A3 Innovation	A4 Consequences for society and environment	A5 Familiarity
√		√	√	√

5.4 Summary

From the above discussion and all the involvement of our research with all those standards and constraints, It is justified that our research is a complex Engineering Problem, and the solution or outcome of our Research is the Complex Engineering Problem solution.

Chapter 6

Conclusion

After all of our study, analysis, and experiments, we've offer a solution to the problem we identified for our research project, and we've surmounted the obstacle of dealing with the problems we encountered during development and research. We have use classification and supervised learning models, Deep Learning models to classify the vulgar speech. After collecting the data sets and pre-process those data and train our models with them. We observed several models and make the comparison among them. We follow the accuracy and loss metric to evaluate our model.

6.1 Limitation

The most difficulty we have faced is finding the exact data set match with our interest. We have collect the data from several online platforms, we also manually generate 1,000 labeled data. Another limitation is the time we got to complete our research.

6.2 Future Work

Many various studies and experiments have been postponed due to a lack of time. Future research will focus on a more in-depth look at specific systems, new proposals to test other methods, or simply curiosity. A number of the suggestions piqued my interest, and I would have liked to give them a try. To make our proposed approach more reliable, we will experiment with more data.

References

[1] Wikipedia. Hate speech — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Hate%20speech&oldid=1038515348, 2021. [Online; accessed 14-August-2021].

[2] DIGITAL IN 2018: world's internet user pass the 4 billion mark. https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018/. Online; Accessed: 2021-10-20.

[3] Cyberbullying in Bangladesh Case Study- how to prevent cyberbullying. https://newsmoor.com/cyberbullying-in-bangladesh-case-study-how-to-prevent-cyberbullying/. Online; Accessed: 2021-10-20.

[4] Wikipedia. List of languages by number of native speakers — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=List%20of%20languages%20by%20number%20of%20native%20speakers&oldid=1035922567, 2021. [Online; accessed 14-August-2021].

[5] Anna Schmidt and Michael Wiegand. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10, Valencia, Spain, April 2017. Association for Computational Linguistics.

[6] Estiak Ahmed Emon, Shihab Rahman, Joti Banarjee, Amit Kumar Das, and Tanni Mittra. A deep learning approach to detect abusive bengali text. In *2019 7th International Conference on Smart Computing Communications (ICSCC)*, pages 1–5, 2019.

[7] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, 2013.

[8] Md. Rezaul Karim, Bharathi Raja Chakravarti, John P. McCrae, and Michael Cochez. Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network. In *7th IEEE International Conference on Data Science and Advanced Analytics (IEEE DSAA,2020)*. IEEE, 2020.

[9] Aiman Snigdha. Aimansnigdha/bangla-abusive-comment-dataset, Jul 2019.

[10] Md Faisal Ahmed. Bangla online comments dataset, Jan 2021.

[11] What is design constraint. https://www.igi-global.com/dictionary/design-of-hexapod-walking-robots/46586#:~:text=1.,features%2C%20and%20the%20robot%20functionality. Online; Accessed: 2021-10-20.

[12] Iea (2015). 25 years washington accord booklet. http://www.ieagreements.org/25_years/25YearsWashingtonAccord-A5booklet-FINAL.pdf. Online; Accessed: 2021-10-20.

[13] Definition of abstract thinking. https://www.healthline.com/health/abstract-thinking. Online; Accessed: 2021-10-20.