

Bengali Paragraph Generation Using Bi-directional LSTM

BY

Sadi Md. Zobaer

ID: 162-15-8052

Naziur Rahman

ID: 162-15-8125

Yusuf Shahriar

ID: 162-15-8140

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Sheikh Abujar

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

Co-Supervised By

Md. Tarek Habib

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JUNE 2020**

APPROVAL

This Project titled “**Bengali Paragraph Generation Using Bi-directional LSTM**”, submitted by Sadi Md. Zobaer, Naziur Rahman, Yusuf Shahriar respectively ID No: 162-15-8052, 162-15-8125, 162-15-8140 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 8 July,2020.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Subhenur Latif

Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

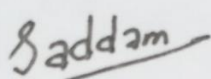


Raja Tariqul Hasan Tusher

Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Saddam Hossain

Assistant Professor

Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

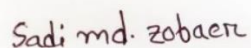
We hereby declare that, this project has been done by us under the supervision of **Sheikh Abujar, Lecturer (Senior Scale)**, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Sheikh Abujar
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:



Sadi Md. Zobaer
ID: 162-15-8052
Department of CSE
Daffodil International University



Naziur Rahman
ID: 162-15-8125
Department of CSE
Daffodil International University



Yusuf Shahriar
ID: 162-15-8140
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Sheikh Abujar, Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of machine learning, natural language processing and data mining to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain, Head, Department of CSE**, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Communication is one of the most important things in our day to day life. Nothing can be done without communication. Each and everything of our life is directly or indirectly dependent on communication. Nowadays communication among peoples crosses the language barrier and all this happens for the rapid growth of technology. It helps us to communicate with people of different languages using our smartphones. All this is possible for the universal use of NLP technology. NLP technology is growing rapidly and trying to make more efficient ways of communication. It's vision is to make a universal way of communication by which all the people of the world can communicate with each other.

In this research, we propose a paragraph generation technique using Bangla language and bi-directional RNN model. here we will generate a paragraph from a given title. from the title we will understand the topic and then we will generate a paragraph based on the title using the prediction of next possible words.

TABLE OF CONTENTS

CONTENTS	PAGE NO
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	2
1.3 Rational of the study	2
1.4 Research Question	3
1.5 Expected Outcome	3
CHAPTER 2: BACKGROUND	4-8
2.1 Introduction	4
2.2 Related Work	4
2.3 Research Summary	7
2.4 Scope of the Problem	7
2.5 Challenges	7
CHAPTER 3: RESEARCH METHODOLOGY	9-18
3.1 Research Subject and Instrumentation	9
3.2 Data collection & pre-processing	13
3.3 Proposed Methodology	14
3.4 Implementation Requirements	18
Chapter 4: Experimental Results and Discussion	19-28
4.1 Introduction	19

4.2 Data segmentation	19
4.3 Data pre-processing	21
4.4 Input file creation	22
4.5 Model building & Dataset fitting	22
4.6 Experiment Accuracy Loss & Output	25
4.7 Summary	28

Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research **29**

5.1 Summary of research	29
5.2 Recommendation	29
5.3 Conclusion	29
5.3 Future work	29

References **30-32**

LIST OF FIGURES

FIGURES	PAGE NO
Fig 2.1: Model loss of text generation using RNN [3]	6
Fig 2.2: Model accuracy of text generation using RNN [3]	6
Fig 2.3: Model Flow chart	8
Fig 3.1: Workflow diagram for paragraph generation.	10
Fig 3.2: Bi-directional LSTM	15
Fig 3.3: Bi-directional LSTM model structure.	17
Fig 4.1: Training data.	23
Fig 4.2: Training data.	24
Fig 4.3: Model accuracy.	25
Fig 4.4: Model loss.	25
Fig 4.5: Output.	26

LIST OF TABLES

TABLES	PAGE NO
Table 3.1: Word numeric assignment for example	11
Table 3.2: Making sentences with sequence	12
Table 3.3: Pad sequencing of sentences	12
Table 3.4: Bengali contraction convert	14
Table 4.1: Data segmentation (news article)	20
Table 4.2: Script file after preprocessing.	21
Table 4.3: Text Output According to Input.	28

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION:

Sequence to sequence text generation is the most proficient methodology for consequently changing over the content of a word from a source succession to an objective succession. Text age is the use of characteristic language age which is helpful in succession demonstrating like machine interpretation, discourse acknowledgment, picture inscribing, language distinguishing proof, video subtitling, and substantially more. Along that Bangla text age has indicated an exceptional job in the field of the individuals whose native language is Bangla [1].

Nowadays technology is one of the most important parts in our day to day life. it makes our life faster and easier. With the rapid growth of technology communication among peoples becomes easier. Even it crosses the language barrier. Now we can communicate with the people of different language and NLP technology is the reason for that. It processes language from one form to another which helps to communicate with people of different language. Nowadays NLP technology is growing much faster and it is trying to make a more efficient way of communication.

Automatic text generation is one of the mentionable exploration regions of Natural Language Processing(NLP). The measure of information is expanding quickly also the need to understand the substance of any content is simply compulsory apparatuses, these days. The region of text generation has been producing for a long time. Mentionable examination has been as of now done through the extractive age approach, on the opposite side-the abstractive text generation, which is the best approach to produce any content like – human. The machine will have the option to give another kind of text, where the comprehension of a given text may be found as like a human-created text. A few exploration improvements have just been accomplished for Abstractive text generation in the English language [2].

its vision is to make an universal way of communication by which all the people of the world can communicate with each other without knowing a common language. Any language having own Unicode can be applied in NLP technology. According to the number of speakers Bengali is the fifth language over the world.in this research we propose a paragraph generation

technique in Bengali language by predicting next possible words.in this we'll generate a paragraph from a given title.

To predict next possible word we will use a word prediction algorithm. In this process firstly the algorithm has to determine the topic on which it's going to predict the next possible words. To do that the algorithm will use dimensional frequency distance of the current and the relevant words. After that the algorithm will analyze the words and will find the word which should be the next word of the current. The most perfect algorithm will predict the most accurate next word and form the most perfect paragraph. As the NLP technology is growing faster and there are many more ways to build a perfect methodology. So we need to work more with the Bangla language and make Bengali language more precious.

1.2 MOTIVATION:

Automation is increasing day by day. People are trying to do things with less efforts. That's why they are inventing and developing almost everything from manual to automatic. In language English is everywhere. But Bengali language also takes a stand in the international field. Day by day Bengali language is getting deeply engaged in every sector of Bengali people. In the literature, education side, print media and many more, Bengali language is advancing. For a thesis or project report or literature review or maybe a long paragraph for anything is tough to write. Automated word and sentence generation in Bengali is done already but there is no paragraph generation for long text. This will help people to fill up long text demands where it is necessary. This will also help Bengali NLP sector where long text paragraphs are needed.

1.3 RATIONALE OF THE STUDY:

There is no uncertainty that there are bunches of shots taken at Natural Language Processing (NLP) in English and these methodologies or the procedures are being utilized in many robotized frameworks just as the mechanical autonomy framework. To grow increasingly mechanized applications or make considerably more proficient in Machine Learning approaches, there is no option in contrast to working with text. This made us intrigued to work with this. Right now, we see that the scratchpad editors are substantially more scholarly. These have a few highlights like auto-rectifications, sentence structure checking, autosuggestion, and so forth. These highlights are the result of the gift of Natural Language Processing. Yet, these

highlights are generally observed for English. Such sort of highlights are uncommon for Bangla text. These, really, take us to work with Text.

1.4 RESEARCH QUESTIONS:

- i.** Would we be able to gather real-time data?
- ii.** Can we apply this in real life?
- iii.** Which are the helpful algorithms those can be utilized for text Demonstrating?
- iv.** Would this research be able to trenchant for the rustic occupants of Bangladesh?

1.5 EXPECTED OUTPUT:

The anticipated result of this research-based project is to assemble an algorithm or make a total productive system that will generate a paragraph from one or more given words as for the built model of the trained dataset.

CHAPTER 2

BACKGROUND

2.1 INTRODUCTION:

In this part, we will portray related works with our project. There are numerous distributed proposal papers for anticipating the exhibition of text generation utilizing diverse models. The summary of those works will be included below. What's more, the venture model, quality and shortcoming of their task will be examined also.

2.2 RELATED WORK:

As all through every one of these years a great deal of work has been done on this field, adding more is expected to add a great value to the workspace. Completely different researchers, various analysts, and enthusiasts are growing new algorithms and procedures just as new models to enhance the current consequences of text generation. Some of them will be talked about in this segment.

Abu Kaisar Mohammad Masum et al. (2019) proposed another kind of text generation of Bangla language utilizing the bi-directional RNN. This method is utilized to foresee the following conceivable word in Bangla content [3]. They basically discussed the n-gram modeling with a pre-trained Bengali word inserting for Text generation and made a Bi-directional Recurrent Neural Network for preparing the model.

Partha Pratim Barman et al. (2018) took a shot at the next word forecast model and in this exploration they utilized RNN based methodology [4]. Fundamentally they constructed a model of LSTM which is a special kind of RNN. They applied this work in a live visiting application where the application can anticipate the following Conceivable word. In this model their objective language was the Assamese language.

Paul Hongsuck Seo et al. (2016) dealt with an inquiry noting framework and utilized the neural system in their framework to consequently anticipate the conceivable answer for the inquiry given by the client [5]. In their conventional neural system (CNN) they utilized a parameter expectation system to make a versatile inquiry noting model. They applied the hashing

procedure to diminish the unpredictability of a huge number of parameters. Accordingly answer determination will be a lot simpler for the framework.

Analyst Ralf Schluter and his group dealt with language modeling utilizing the LSTM neural system in 2012 [6]. The fundamental reason for their exploration was to manufacture a neural system that can expand the precision of the investigation. They utilized their model on the English language and just as an enormous arrangement of French language demonstrating the task. They have demonstrated improvements in perplexity over standard repetitive neural system Language models.

Wen-tau Yih and his group from Microsoft Research chipped away at linguistic regularities [7]. In this exploration they attempted to assemble a model that can foresee the space word portrayal for an inquiry noting framework. This framework incorporates vector portrayal and can naturally get familiar with the relationship of words. With about 40% exactness on symmetric inquiry replying. They illustrate that the word vectors catch semantic regularities by utilizing the vector balance technique. Their work was the best model in those days.

Researcher Md. Sanzidul Islam and his team stripped away LSTM recurrent neural network [8]. In their expedition they generated Bangla sentences using sequence to sequence sentence generation technique. They have shown their expertise in deep learning approach, Long-short Term Memory (LSTM), a special kind of RNN. They analyzed the sequence of text data and predicted the adjacent word. They suggested a simulated Bangla Text Generator which is one of very few of its kind. They have accomplished a very adequate accuracy rate.

A character-based encoder-decoder model which is obtained to transliterate sequence to sequence by Amir H. Javidinejad [9]. They proposed an encoder worked with Bidirectional Recurrent Neural Network that encodes a succession of mantels into vector representations with fixed length.

Researcher Zhang and Lapata clarifies momentous work utilizing Recurrent Neural Network (RNN) [10]. They have shown tremendous performance over RNN using Chinese poetry. Their initiative was very decent that their model could recreate some lines of Chinese poems automatically. It was a very extraordinary work then.

Praveen Krishnan and accompanying enthusiasts presented an OCR framework which seeks after a joined architecture in seven distinct dialects of India and a division free technique [11]. Their framework was proposed to help the consistent learning in the hour of being it usable, as constant client input. They worked with Bi-directional Long Short-term Memory (BLSTM) strategy, another type of general LSTM.

Robert Ostling et al. suggested a model that was built with sequence-to-sequence fictitious neural systems and LSTM engineering that was a major regard for enthusiasts [12].

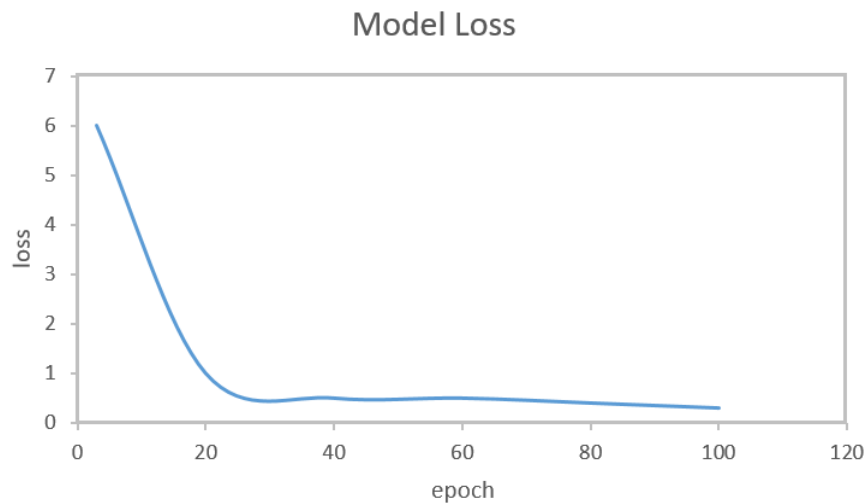


Fig 2.1: Model loss of text generation using RNN [3]

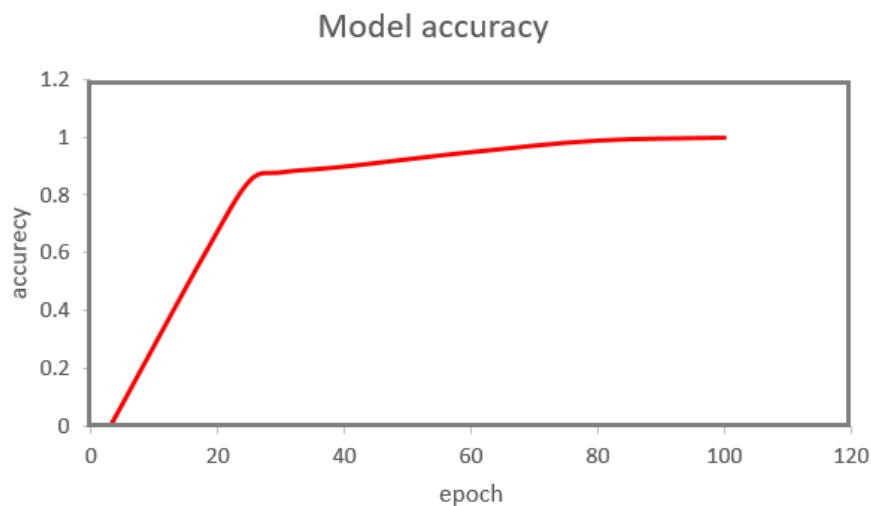


Fig 2.2: Model accuracy of text generation using RNN [3]

2.3 RESEARCH SUMMARY:

So it has been seen that a great deal of research has just been done on Text generation for the English language, etc. for the Bengali language by arrangement of sequence to sequence word or text prediction, and different data mining approaches. The majority of them utilized the greatest measure of foundation data as training data and the remainder of the data as test data. What's more, the two sorts of information yield a definitive outcome and are joined with the assistance of the applied classifier model.

2.4 SCOPE OF THE PROBLEM:

At present we are taking advantage of innovation and mechanization frameworks a great deal we are accustomed to having computerized recommendations of a wide range of text-based things, for example, auto sentence proposal, auto-remedy of words, also in our speech we want to have a suggestion, for example, commanding a system. There have been numerous works upon auto sentence and word generation yet exceptionally less work in computerized paragraph generation in the English language and there is no work at present about a paragraph generation system or approach in the Bengali language. By the assistance of this model, we can have an auto-generated paragraph from a given word and furthermore from a succession of given words in Bengali it can help us in a wide range of areas, for example, instructive reason, print media, writing, web-based social networking, an online blog and any kind of website working with Bangla and require paragraph generation feature can take advantage of this research project.

2.5 CHALLENGES:

Collecting Data

Data collection was quite easy. From various online Bengali news portal, Facebook and various Bengali blog posts, we got our data. Most of the data is from news. Also manually we put some text.

Language friendly system making

The troublesome part was making words and sentences good with the system. In any case, by taking a few aides from online it gets simpler. We found and used Bengali stopwords from the internet. We found 393 stopwords. Every one of these things made our project kind of difficult. But when there is a problem, there is a solution too. It made us learn many new things.

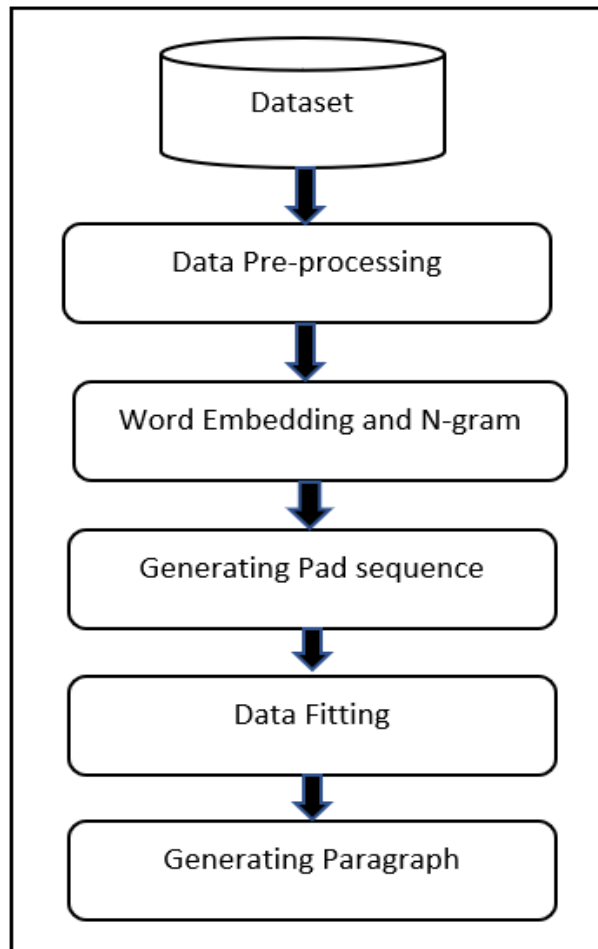


Fig 2.3: Model Flow chart

Selecting model

Determining a model is kind of a tough job. If we choose the wrong model then the whole project would be devastating and a loss of valuable time. We used some layers to create a model. Bi-Directional LSTM, Batch Normalization, Dropout, Dense, Activation. We used these to provide the best result to show an output of an automated paragraph regarding the input being either one or more words.

Method of project

The proposed procedure of paragraph generation on word or sentence which is appeared in Figure 2.3. It has secured the means of text to paragraph and it is consisting basic six steps.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 RESEARCH SUBJECT AND INSTRUMENTATION:

Language Modelling is the most momentous piece of modern NLP. There are some pieces of the assignment, for example, Machine Interpretation, Text Generation, Text Summarization Speech to Text Generation, Text to Speech and so on. A good language model acquires information from past arrangements of words. In this paper we worked with N-gram token sequences, Word Embedding and Bi-direction LSTM to generate a paragraph from a given title using next word prediction.

Firstly, we collect our Dataset and then we clear our data for training purpose and then we build our corpus for training purpose. we use N-gram token sequence after that we apply word embedding. we use bn_w2v_model.txt for word embedding. After that we created our model and train out corpus with the corpus and at last we generate paragraph from given title. Here is the figure of our workflow.in this figure we have shown our working procedure step by step.

N-gram:

N-gram is used in NLP base research works. N-gram is 'n' types of sequence using word, para, sentences etc. It means when 'n' types of items continuously get sequenced from a given corpus. A language model needs a setup of the token and which can predict the next word or grouping. So we need to tokenize the words. We use Keras work in a tokenize model which condenses words with their record number from the corpus and after that all content changes the arrangement of the token.

The arrangement contains tokens which was produced using the corpus in N-gram. Each token tells to the chronicle the word which is in the corpus vocabulary. N-grams can be used as a uni-gram, bi-gram or tri-gram model depending on the dataset and the goal of that research.

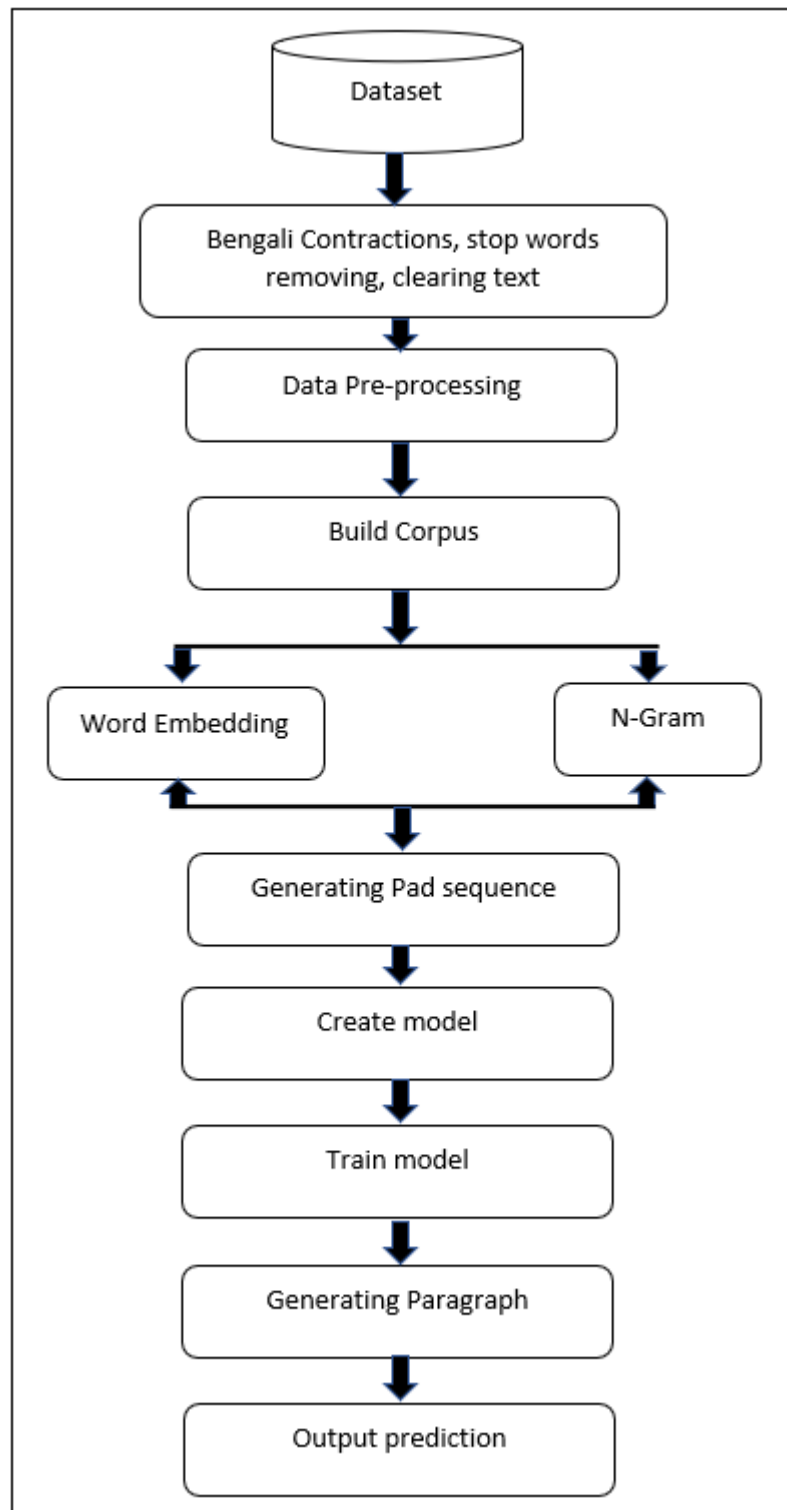


Fig 3.1: Workflow diagram for paragraph generation.

Word Embedding:

To feed the data to machine word embedding is needed. The machine and algorithm only understands numerical data. So word embedding is converting text/word into vectors of real numbers. There are two word embedding approaches. Context co-occurrences and predict words in context. Word2vec is used to generate word embedding. There are some reasons why we used word embedding. A machine can understand just paired or numerical honor. So, processing a language and working with languages must be required word2vector.

Every vector speaks to the vocabulary of the corpus document when applying word extending each machine can change over tokenize word to a vector . When all of the word now has a vector value. Now word embedding is to find after which value will be which other value. That means After which word will be which word There are several Word embedding pre-trained models in different languages but in the Bengali language few numbers of word embedding files present, most of which is not enough for research. One good pre-trained model was found which is usable and we used it in our research purposes. Which name is “bn_w2v_model.txt”.

Pad sequences:

As known, because for ML models for NLP tasks it must be feeded matrix of normalized values, that means all sentences (tokens) must have the same number of elements, even if there are blanks or none. So, padding is needed for long or short text to have the same size. Some text input can be short, some can be long. All the neural networks work with the same shape and size inputs. So when a sentence is too short padding adds a '0' by default and when it is too long then the first value/words gets cut.

Word	Numeric Assignment
I	5
am	2
a	4
boy	3
play	6
football	9
everyday	8
strong	10

Table 3.1: Word numeric assignment for example

‘ I am a boy ’

‘ I play football everyday ’

‘ I am strong ’

When padding is used, every word gets a numeric value to identify them. Suppose, I=5, am=2, a=4, boy=3, play=6, football=9, everyday=8, strong=10.

So, numerically [[5, 2, 4, 3], [5, 6, 9, 8], [5, 2, 10]] this represents all the sentences above.

Sentences	Sequence
I am a boy	[5, 2, 4, 3]
I play football everyday	[5, 6, 9, 8]
I am strong	[5, 2, 10]

Table 3.2: Making sentences with sequence

So there are 10 unique words. When padding length will be fixed like 4 then all the sentences will be 4 words. After defining padding max length 4 the result will be

[[5, 2, 4, 3], [5, 6, 9, 8], [0, 5, 2, 10]]

Sentences	Sequence
I am a boy	[5, 2, 4, 3]
I play football everyday	[5, 6, 9, 8]
{Blank} I am strong	[0, 5, 2, 10]

Table 3.3: Pad sequencing of sentences

Here ‘0’ is OOV or out of vocabulary which indicates a blank text. But if the padding length is 3 then the result will be [[2, 4, 3], [6, 9, 8], [5, 2, 10]].

This is how padding works.

Every development has an alternative length. So we need to pad sequences for making an arrangement of balanced length. For this point, we use Keras pad sequences function. The undertaking of the learning of the model we used n-gram gathering from a given word and the anticipated word as the concurrent word.

3.2 DATA COLLECTION & PREPROCESSING:

Each profound learning algorithm needs countless data. As enormous the dataset is the outcome is that vastly improved. For our model, we need an attractive number of data as well. In any case there isn't sufficient dataset accessible on the web, along these lines, we have gathered different kinds of data, for example, news, Facebook posts, and so on and afterward made a synopsis of those writings for our research reason. In our dataset, there are just two segments accessible which is important and those are text portrayals also, outlines. We collected data from:

- From news data
- From blog data

Most of the data of our dataset is collected from news data as there are a lot of sources of news data. About 90 percent of our data is collected from news data and the rest 10 percent of data is collected from blog posts from many blog sites.

After the collection of data there comes the pre-processing of the data. We pre-processed our data to be used for training purposes. To pre-process our data, we used three techniques. They are:

- Change contractions
- Remove stopwords
- Clean text

Now we will tell details about these three techniques of data pre-processing. After finishing all these three we will find a clean corpus that will be used for training purposes.

Change contractions:

Before getting a ready dataset for content age, we have to include Bengali compressions. Since contraction contains a short type of a word, firstly we will change contractions with the full form. It will make our data cleaner to be used in training, so we need to perform data contraction to have a more accurate text corpus. After having this in our text our data is somewhat processed already for training purposes. Here we show some examples of data contractions.

Contractions	Full Form
রেজিঃ	রেজিস্ট্রেশন
ডাঃ	ডাক্তার
মোঃ	মোহাম্মদ
ইঞ্জিঃ	ইঞ্জিনিয়ার
বিঃদ্রঃ	বিশেষ দ্রষ্টব্য

Table 3.4: Bengali contraction convert

Remove stopwords:

Stop words are words that are commonly used in any language which has no actual meaning. It's just used to make a sentence full of sweetness but it doesn't carry any tangible meaning. So, we have to remove the stop words from the text. To remove stop words in our paper we used remove stopwords function. Removing stop words removes the redundancy of sentences. After removing the stop words, we will find a more efficient dataset that will be used for training purposes.

Clean text:

Our collected data is not as much usable to train our model with this data so we need to clean our data first. To clean our data, we have to remove extra whitespaces, we need to remove digits, we need to remove special characters and so on. To clear our data properly we used some regular expressions. to clean data some regular expressions were used. We used all regular expressions in a function and called it where needed. After cleaning our data our corpus is ready for further use like training purpose.

3.3 PROPOSED METHODOLOGY:

In this research we worked Bi-directional LSTM which is on kind of special RNN. it is one of the most popular layers for text base auto completion. it works in two directions one is forward and another is backward. they are in different position. Forward direction gives us the predicting output and on the other hand Backward direction stores the previous information's and outputs.

The formula will be,

$$hf = \sigma(Wf * X + hf + bf) \quad (1)$$

$$hb = \sigma(Wb * X + hb + bb) \quad (2)$$

$$y = (hfWf + hbWb + b) \quad (3)$$

Here,

σ = Activation Function

hf = Forward hidden layer

hb = Backward hidden layer

W = Weight,

B = Bias

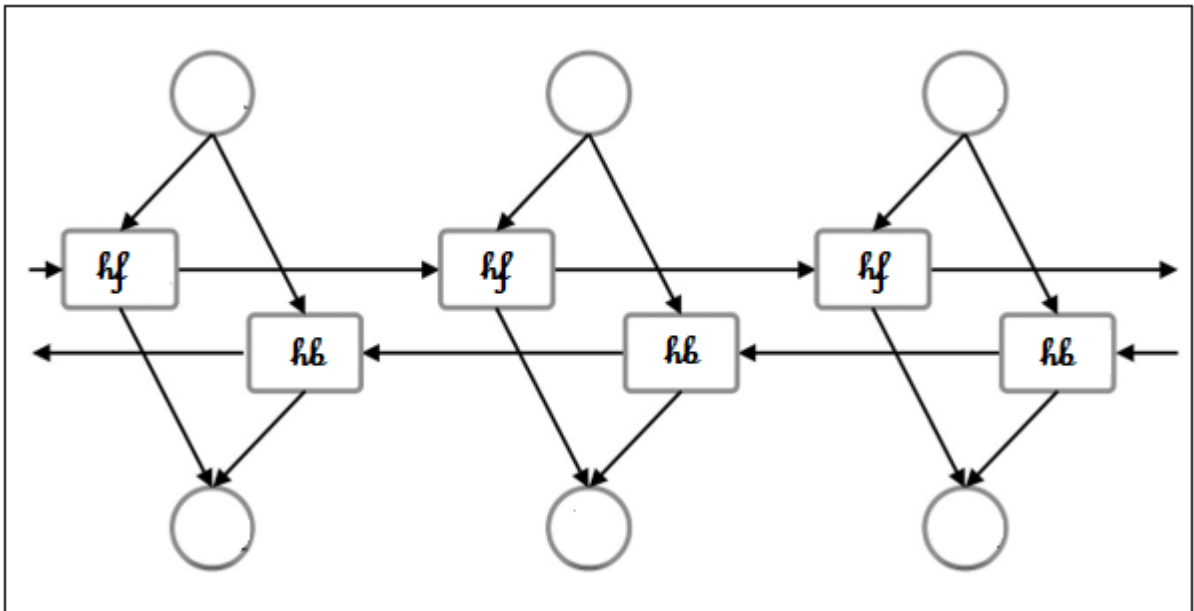


Fig 3.2: Bi-directional LSTM

As we know that bidirectional LSTM stores previous outputs and information and this previous information are used in current state as input. so it predicts the next remembering the previous. in our research text sequences and time is used as input. For training our model we used a sequential model of keras. in this sequential model we used LSTM layer inside bidirectional layer and used 516 units and 'ReLU' as activation function for this layer. As LSTM is a special kind of RNN in cancels angle. Every LSTM cell has three entry points. Input Gate, Forget Gate, Output Gate and a cell state included data across the doors.

$$it = \sigma(wi[ht-1,xt] + bi) \quad (4)$$

$$ft = \sigma(wf[ht-1,xt] + bf) \quad (5)$$

$$ot = \sigma(w_o[ht-1,xt] + b_o) \quad (6)$$

$$ct = ft * ct-1 + it * \sigma(w_c[ht-1,xc] + b_c) \quad (7)$$

$$ht = ot * \sigma(ct) \quad (8)$$

Here,

it = input gate's,

ft = forget gate's ,

ot = output gate,

ct = cell state,

ht = hidden state,

σ = activation function

After that we used Batch Normalization layer to normalize our trainable data. After that we used Dropout layer with the 0.4 value and it helps us to reduce overfitting of data. After that we used a Dense layer with the activation function 'Softmax'. In the Dense layer we passed total words as a parameter and again used the Batch Normalization layer to normalize our data. It also reduces are trainable parameters.to reduce the model loss we expand the learning rate from 0.0001 to 0.001 and to reduce the time of every epoch we used batch size 100. Finally, we fit the defined model and set input and output sequence with verbose.

Algorithm used in our model is:

1. **Function** model create (max sequence length, total words):
2. input length= max sequence length-1
3. declare **Sequential** ()
4. add(**Bidirectional(LSTM)(units),input shape**)
5. add(**BatchNormalization()**)
6. add(**Activation**)
7. add(**Dropout(0.4)**)
8. add(**Dense(total words)**)
9. add(**BatchNormalization()**)
10. add(**Activation**)
11. compile(loss function, optimizer)
12. **return** model
13. create model(max sequence length, total words)

This is the pseudo code for our model. We used Keras sequential models. In this model we used bidirectional LSTM as input layer after that we normalized using Batch Normalization and then we used activation layer and passed result as activation function. After that we used Dropout layer with a value of 0.5 to prevent overfitting. After that we used a dense layer with Batch Normalization layer. We also increased learning rate to reduce loss and used batch size 100 to reduce training time for each epoch. We used 100 epochs for our research purpose. We fit 600 news data on our model and found the accuracy of 96.33%.

Here we show the model structure of bi directional LSTM.

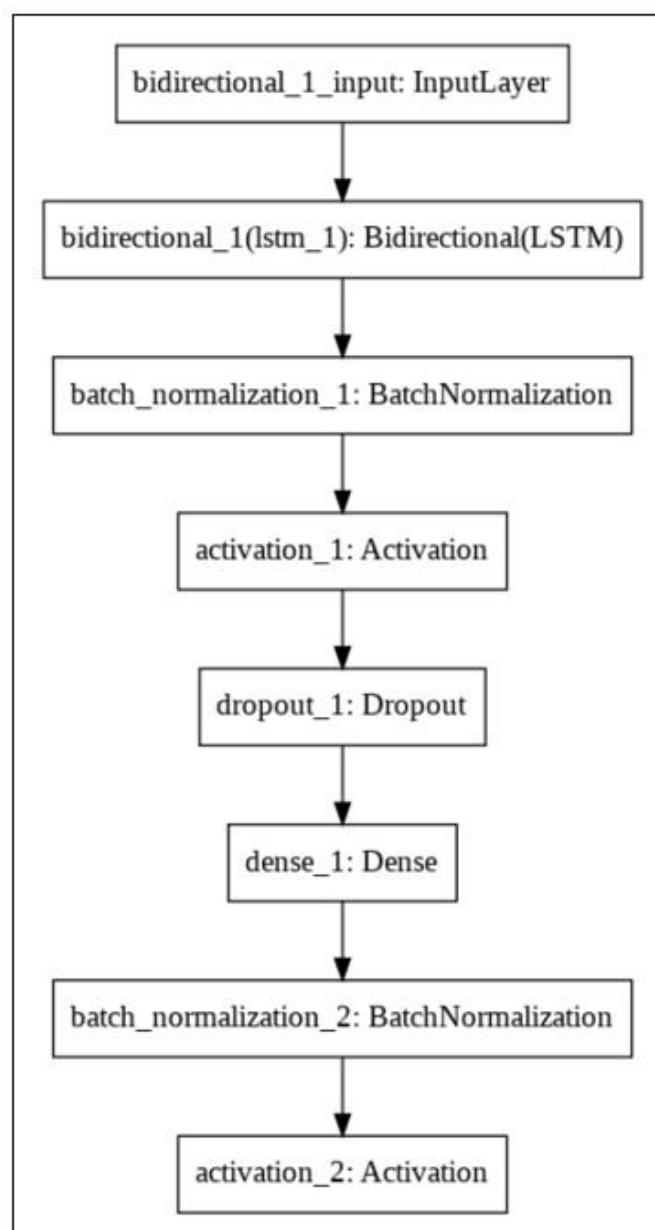


Fig 3.3: Bi-directional LSTM model structure

3.4 IMPLEMENTATION REQUIREMENTS:

We needed some hardware and software to train our data on our research model and feed the data to the machine. We are discussing them below. Anaconda is a free and open source Python distribution and R programming languages. It works for predictive analytics, Data science, Machine learning, Image processing and scientific computing. It simplifies development and management of a project. Package management system named conda manages the package version. Anaconda navigator provides us with various development environments to use like spyder, jupyter notebook, orange and so on. all we have to do is to launch it from anaconda navigator and use it. Python is a high level and most popular interpreted dynamic programming language. It becomes more popular because it emphasizes code reliability.it provides us a huge set of libraries, packages, methods to work with. when we write code with python it's like we are commanding someone to do it. It provides automatic memory management. Multiple programming paradigms are supported by Python.it has a huge amount of libraries for data science, machine learning, data analysis, image processing, data mining etc. It is best for research based work because it simplifies the code. Colaboratory or “Colab” is a product from Google which allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs. It's very useful for students who are very interested in learning Ai, Machine learning, Data science and so on but don't have that configuration pc to train models with. Colabratory provides us tensor flow gpu for 12 hour. So we can easily train our model without having high configuration pc. So it's very useful for us.

Hardware requirements:

1. Operating system (Windows 7 and above)
2. Web browser (Chrome or Firefox)
3. Hard disk (512 GB)
4. Ram (8 GB)

Chapter 4

Experimental Results and Discussion

4.1 INTRODUCTION

Regarding paragraph generation, we have attempted to build up an algorithm by which one can consequently generate a paragraph from at least one given word. In section 3, we have depicted the research project and the technique of its working. Presently, in section 4, we have attempted to show the yield results outcome and procedure of the paragraph generation.

4.2 DATA SEGMENTATION

For this research project, we gathered preprocessed data from prestigious newspapers and online articles. Yet at the same time, there is some repetitive data there. The articles are put away on a text document file. At that point, the information is changed into an excel document record and fragmented into two areas named title and paragraph. A perspective on portioned data is given below.

Title	Paragraph
নিজামী	রায়ে মতিউর রহমান নিজামীকে দোষী সাব্যস্ত করতে তৎকালীন শিল্পসচিব শোয়েব আহমদ ও বাংলাদেশ রসায়ন শিল্প সংস্থার (বিসিআইসি) চেয়ারম্যান মেজর জেনারেল (অব.) ইমামুজ্জামানের বক্তব্য আমলে নিয়েছেন আদালত। শিল্পমন্ত্রী হিসেবে সিইউএফএল জেটিঘাট নিজামীর অধীন স্থাপনা ছিল এবং এটা ছিল বিশেষ নিরাপত্তা স্থাপনা। সেখানে অস্ত্র খালাস ও আটকের বিষয়ে নিজামীকে জানিয়েছিলেন এই দুজন শীর্ষস্থানীয় কর্মকর্তা। তাঁরা আদালতকে বলেছেন, তখন নিজামী তাঁদের বলেছিলেন, ‘আমি বিষয়টা শুনেছি। এটা আইনশৃঙ্খলা বাহিনী দেখছে। দেশের হাইয়েস্ট অথরিটিও (সর্বোচ্চ কর্তৃপক্ষ) এ বিষয়ে অবগত আছেন। স্বরাষ্ট্র মন্ত্রণালয় এ ব্যাপারে পদক্ষেপ নিচ্ছে। কাজেই আমাদের কোনো কিছু করার প্রয়োজন নেই।’ আসামি ও সাক্ষীদের বক্তব্যের উদ্ধৃতি দিয়ে বলা হয়, বিশাল চালানের অস্ত্র ও গোলাবারুদ নিরাপদে সিইউএফএল জেটিঘাটে খালাসের জন্য তৎকালীন ভারপ্রাপ্ত শিল্পসচিব নুরুল আমিনকে সিইউএফএল রেস্টহাউসে পাঠিয়েছিলেন তখনকার শিল্পমন্ত্রী নিজামী। এসব বিবেচনায় নিয়ে আদালত তাঁকে দোষী সাব্যস্ত করেন।
বাবর	লুৎফুজ্জামান বাবরের সম্পৃক্ততার তথ্য আদালতে বর্ণনা করেন ২৪ জন সাক্ষী ও আসামি। আদালত ডিজিএফআইয়ের সাবেক মহাপরিচালক মেজর জেনারেল (অব.) সাদিক হাসান রুমী ও তৎকালীন অধিনায়ক কর্নেল এ কে এম রেজাউর রহমান, পুলিশের তৎকালীন উপকমিশনার আবদুল্লাহেল বাকী, তৎকালীন চট্টগ্রাম মহানগর পুলিশ কমিশনার এস এম সাকিবর আলী ও সাবেক স্বরাষ্ট্রসচিব ওমর ফারুককে বক্তব্য আমলে নেন। তদন্তের সময় অস্ত্রগুলো এনএসআইয়ের কর্মকর্তাদের তত্ত্বাবধানে খালাসের বিষয়টি এড়িয়ে যাওয়ার জন্য বাবর নির্দেশ দিয়েছিলেন। ফলে বিশাল চালানের অস্ত্র খালাসের আসল তদন্ত ধামাচাপা পড়ে যায়। আদালত সূত্র জানায়, বাবরকে দোষী সাব্যস্ত করার ক্ষেত্রে এটি অন্যতম ভূমিকা রেখেছে। রেজ্জাকুল হায়দার: মেজর জেনারেল (অব.) রেজ্জাকুল হায়দার চৌধুরীকে দোষী সাব্যস্ত করা হয় আরেক আসামি উইং কমান্ডার সাহাব উদ্দিন ও ডিজিএফআইয়ের

	<p>তৎকালীন মহাপরিচালক মেজর জেনারেল (অব.) সাদিক হাসান রুমীর জবানবন্দির ভিত্তিতে। ২০০৪ সালের মার্চে ঢাকা সি.এম.এইচ. চিকিৎসাধীন থাকা অবস্থায় সাহাব উদ্দিনকে দেখতে যান উলফা নেতা পরেশ বড়ুয়া। পরেশ বড়ুয়াকে বিদায় দেওয়ার সময় দরজা খুলে সাহাব উদ্দিন দেখতে পান রেজ্জাকুল হায়দারকে। এ ছাড়া ডিজিএফআইয়ের তৎকালীন মহাপরিচালক সাদিক হাসান রুমী ২০০৪ সালের ২ এপ্রিল বাবরের সঙ্গে অস্ত্র দেখার জন্য চট্টগ্রামে আসেন। তখনই ডিজিএফআইয়ের অধিনায়ক কর্নেল এ কে এম রেজাউর রহমান অস্ত্র চোরাচালানের সঙ্গে রেজ্জাকুল হায়দারের জড়িত থাকার বিষয়টি সাদিক হাসানকে জানিয়েছিলেন বলে তাঁদের জবানবন্দিতে জানা যায়। আদালতে দেওয়া তাঁদের বক্তব্য আমলে নিয়ে রেজ্জাকুল হায়দারকে দোষী সাব্যস্ত করা হয় বলে সূত্র জানিয়েছে।</p>
আবদুর রহিম	<p>এনএসআইয়ের মহাপরিচালক ব্রিগেডিয়ার জেনারেল (অব.) আবদুর রহিম প্রসঙ্গে বলা হয়, পাকিস্তানের মালিকানাধীন দুবাইভিত্তিক এআরওয়াই গ্রুপের আমন্ত্রণে ২০০৩ সালে ব্রিগেডিয়ার আবদুর রহিম সন্ত্রাসীক দুবাইসফরে গিয়ে পাকিস্তান গোয়েন্দা সংস্থা প্রধানের সঙ্গে বৈঠক করেন। একই বছর এআরওয়াই গ্রুপ রহিমকে ৪০ হাজার পাউন্ড উপহার দেয়। ঢাকার সেফ হাউসে পাকিস্তানের হাইকমিশনারকে নিয়ে পরেশ বড়ুয়াসহ একাধিক বৈঠক করেন রহিম। ২০০৪ সালের মার্চে মহাপরিচালক রহিম তাঁর কক্ষে উইং কমান্ডার সাহাব উদ্দিন ও মেজর লিয়াকতকে ডেকে পাঠান। মহাপরিচালক তাঁদের বলেন, এআরওয়াই গ্রুপের অর্থায়নে ওই দেশের সহায়তায় কিছু অস্ত্র ও যন্ত্রপাতি আসবে। এ থেকে আদালত নিশ্চিত হন, ঘটনার সঙ্গে রহিম শুরু থেকেই জড়িত ছিলেন এবং সে জন্য তাঁকে দোষী সাব্যস্ত করা হয়।</p>
সাহাব উদ্দিন	<p>এনএসআইয়ের পরিচালক উইং কমান্ডার (অব.) সাহাব উদ্দিন আহাম্মদ দোষী সাব্যস্ত হন মাঠ কর্মকর্তা আকবর হোসেন খান এবং মহাপরিচালক আবদুর রহিমের বক্তব্যের ভিত্তিতে। মাঠ কর্মকর্তা আকবর হোসেন সাহাব উদ্দিনের নির্দেশে অস্ত্র পরিবহনের জন্য চট্টগ্রামের গ্রিনহাউস ট্রান্সপোর্ট থেকে ট্রাক ভাড়া করেন। অস্ত্র আটকের পর সাহাব উদ্দিনকে বিমানবাহিনীতে বদলি করে পাঠানো হলে আমেরিকান দূতবাসের দুই কর্মকর্তা দিয়ে তিনি রহিমের কাছে তদবির করান। আকবর ও রহিম তা আদালতে স্বীকারোক্তি মূলক জবানবন্দিতে উল্লেখ করেন।</p>
লিয়াকত হোসেন	<p>এনএসআইয়ের উপপরিচালক মেজর (অব.) লিয়াকত হোসেনকে দোষী সাব্যস্ত করা হয় তৎকালীন পুলিশ কমিশনার সাকিব আলী ও উপকমিশনার আবদুল্লাহেল বাকীসহ আরও কয়েকজন পুলিশ সদস্যের জবানবন্দির ভিত্তিতে। আদালত সূত্র জানায়, ঘটনার দিন ২০০৪ সালের ১ এপ্রিল দিবাগত রাত আড়াইটা-তিনটার দিকে উপকমিশনার (বন্দর) আবদুল্লাহেল বাকী পুলিশ কমিশনার সাকিব আলীকে ফোনে জানান, অবৈধ মালামালের একটি বাক্স খুলে অস্ত্র ও গোলাবারুদ দেখা গেছে। ঘটনাস্থল থেকে পাঁচজনকে আটক, পাঁচটি ট্রাক ও দুটি ট্রলার জব্দ করা হয়। ঘটনাস্থলে উপস্থিত ব্যক্তিদের মধ্যে একজন নিজেকে এনএসআইয়ের উপপরিচালক মেজর লিয়াকত বলে পরিচয় দিয়ে উর্ধ্বতন কর্তৃপক্ষের নির্দেশে মালামাল খালাস করা হচ্ছে জানিয়েছিলেন। বাধা দেওয়া হলে লিয়াকত আবদুল্লাহেলকে বলেছিলেন, ‘আপনি কেন এ মালামাল আটক করছেন। এ মালামাল খালাসের ব্যাপারে সহায়তা করুন।’</p>

Table 4.1: Data segmentation (news article)

In the Fig 4.1, we can see a portion of the articles are covered with HTML labels. They are the genuine sectioned information containing immense measures of stop words, data contractions, etc. So information preprocessing is expected to lessen information redundancy to locate the specific information or data. So, we cannot input the data into our model. We must convert our data into the face of machine understanding format which we can feed then to the model.

4.3 DATA PRE-PROCESSING

We have structured a Python content script to start an information pre-preparing task. The content is fundamentally working for:

- Removing unwanted characters.
- Getting rid of unnecessary blank spaces.
- Removing new lines.
- Removing stopwords.
- Expanding contraction words.

Paragraph: 1	<p>রায়ে মতিউর রহমান নিজামীকে দোষী সাব্যস্ত করতে তৎকালীন শিল্পসচিব শোয়েব আহমদ ও বাংলাদেশ রসায়ন শিল্প সংস্থার বিসিআইসি চেয়ারম্যান মেজর জেনারেল অব ইমামুজ্জামানের বক্তব্য আমলে নিয়েছেন আদালত শিল্পমন্ত্রী হিসেবে সিইউএফএল জেটিঘাট নিজামীর অধীন স্থাপনা ছিল এবং এটা ছিল বিশেষ নিরাপত্তা স্থাপনা সেখানে অস্ত্র খালাস ও আটকের বিষয়ে নিজামীকে জানিয়েছিলেন এই দুজ</p>
Title: নিজামী	
Paragraph	
Paragraph: 2	<p>লুৎফুজ্জামান বাবরের সম্পৃক্ততার তথ্য আদালতে বর্ণনা করেন জন সাক্ষী ও আসামি আদালত ডিজিএফআইয়ের সাবেক মহাপরিচালক মেজর জেনারেল অব সাদিক হাসান রুমী ও তৎকালীন অধিনায়ক কর্নেল এ কে এম রেজাউর রহমান পুলিশের তৎকালীন উপকমিশনার আবদুল্লাহেল বাকী তৎকালীন চট্টগ্রাম মহানগর পুলিশ কমিশনার এস এম সাক্বির আলী ও সাবেক স্বরাষ্ট্রসচিব ওমর ফারুকের বক্তব্য আমলে নেন তদন্তের সময়</p>
Title: বাবর	
Paragraph	
Paragraph: 3	<p>এনএসআইয়ের মহাপরিচালক ব্রিগেডিয়ার জেনারেল অব আবদুর রহিম প্রসঙ্গে বলা হয় পাকিস্তানের মালিকানাধীন দুবাইভিত্তিক এআরওয়াই গ্রুপের আমন্ত্রণে সালে ব্রিগেডিয়ার আবদুর রহিম সস্ত্রীক দুবাই সফরে গিয়ে পাকিস্তান গোয়েন্দা সংস্থা প্রধানের সঙ্গে বৈঠক করেন একই বছর এআরওয়াই গ্রুপ রহিমকে হাজার পাউন্ড উপহার দেয়</p>
Title: আবদুর রহিম	
Paragraph	
Paragraph: 4	<p>এনএসআইয়ের পরিচালক উইং কমান্ডার অব সাহাব উদ্দিন আহম্মদ দোষী সাব্যস্ত হন মাঠ কর্মকর্তা আকবর হোসেন খান এবং মহাপরিচালক আবদুর রহিমের বক্তব্যের ভিত্তিতে মাঠ কর্মকর্তা আকবর হোসেন সাহাব উদ্দিনের নির্দেশে অস্ত্র পরিবহনের জন্য চট্টগ্রামের গ্রিনহাউস ট্রান্সপোর্ট থেকে ট্রাক ভাড়া করেন</p>
Title: সাহাব উদ্দিন	
Paragraph	

Table 4.2: Script file after preprocessing.

4.4 INPUT FILE CREATION

After a productive preprocessing process, we have some absolute news records close by. At that point, we need to perform Natural Language Process on the news documents, we should join every one of these records into a record. For this, we utilize another python content. This record takes the organizer name that contains all .py documents as an information and delivers just a document where all news contained separately being blended and split into sentence segments of length 10.

4.5 MODEL BUILDING & DATASET FITTING

Preprocessed data then spited into 10-gram segments to fit into our built model.

- Number fitted paragraph- 600

After building the corpus we passed that into our model and to decrease the loss of the model we have increased the learning rate of the optimizer. As we used Adam as an optimizer its default learning rate was 0.0001 and we passed the learning rate of 0.001 into the model compilation. To reduce the time of each epoch we increased the batch size in model compilation. It reduces our training time in every epoch to work the model more efficiently.


```

Epoch 50/100
- 53s - loss: 0.4632 - accuracy: 0.9237
Epoch 51/100
- 53s - loss: 0.4422 - accuracy: 0.9282
Epoch 52/100
- 53s - loss: 0.4336 - accuracy: 0.9302
Epoch 53/100
- 53s - loss: 0.4257 - accuracy: 0.9309
Epoch 54/100
- 53s - loss: 0.4160 - accuracy: 0.9328
Epoch 55/100
- 53s - loss: 0.4023 - accuracy: 0.9342
Epoch 56/100
- 53s - loss: 0.3916 - accuracy: 0.9371
Epoch 57/100
- 53s - loss: 0.3878 - accuracy: 0.9371
Epoch 58/100
- 54s - loss: 0.3842 - accuracy: 0.9380
Epoch 59/100
- 53s - loss: 0.3680 - accuracy: 0.9404
Epoch 60/100
- 53s - loss: 0.3640 - accuracy: 0.9405
Epoch 61/100
- 53s - loss: 0.3574 - accuracy: 0.9420
Epoch 62/100
- 53s - loss: 0.3478 - accuracy: 0.9443
Epoch 63/100
- 52s - loss: 0.3435 - accuracy: 0.9447
Epoch 64/100
- 53s - loss: 0.3375 - accuracy: 0.9455
Epoch 65/100
- 52s - loss: 0.3307 - accuracy: 0.9471
Epoch 66/100
- 52s - loss: 0.3289 - accuracy: 0.9468
Epoch 67/100
- 52s - loss: 0.3237 - accuracy: 0.9482
Epoch 68/100
- 52s - loss: 0.3177 - accuracy: 0.9496
Epoch 69/100
- 52s - loss: 0.3119 - accuracy: 0.9498
Epoch 70/100
- 53s - loss: 0.3077 - accuracy: 0.9507
Epoch 71/100
- 52s - loss: 0.3020 - accuracy: 0.9518
Epoch 72/100
- 52s - loss: 0.2977 - accuracy: 0.9528
Epoch 73/100
- 52s - loss: 0.2945 - accuracy: 0.9532
Epoch 74/100
- 53s - loss: 0.2895 - accuracy: 0.9536
Epoch 75/100
- 53s - loss: 0.2881 - accuracy: 0.9544

```

Fig 4.1: Training data.

```
Epoch 76/100
- 55s - loss: 0.2861 - accuracy: 0.9541
Epoch 77/100
- 56s - loss: 0.2814 - accuracy: 0.9545
Epoch 78/100
- 54s - loss: 0.2750 - accuracy: 0.9556
Epoch 79/100
- 53s - loss: 0.2756 - accuracy: 0.9554
Epoch 80/100
- 53s - loss: 0.2728 - accuracy: 0.9563
Epoch 81/100
- 54s - loss: 0.2659 - accuracy: 0.9581
Epoch 82/100
- 53s - loss: 0.2666 - accuracy: 0.9571
Epoch 83/100
- 53s - loss: 0.2575 - accuracy: 0.9587
Epoch 84/100
- 53s - loss: 0.2590 - accuracy: 0.9583
Epoch 85/100
- 54s - loss: 0.2599 - accuracy: 0.9584
Epoch 86/100
- 54s - loss: 0.2550 - accuracy: 0.9599
Epoch 87/100
- 55s - loss: 0.2554 - accuracy: 0.9583
Epoch 88/100
- 54s - loss: 0.2520 - accuracy: 0.9596
Epoch 89/100
- 54s - loss: 0.2474 - accuracy: 0.9610
Epoch 90/100
- 54s - loss: 0.2482 - accuracy: 0.9601
Epoch 91/100
- 55s - loss: 0.2451 - accuracy: 0.9613
Epoch 92/100
- 55s - loss: 0.2428 - accuracy: 0.9613
Epoch 93/100
- 55s - loss: 0.2405 - accuracy: 0.9621
Epoch 94/100
- 55s - loss: 0.2382 - accuracy: 0.9626
Epoch 95/100
- 54s - loss: 0.2392 - accuracy: 0.9624
Epoch 96/100
- 54s - loss: 0.2360 - accuracy: 0.9629
Epoch 97/100
- 54s - loss: 0.2321 - accuracy: 0.9631
Epoch 98/100
- 54s - loss: 0.2339 - accuracy: 0.9626
Epoch 99/100
- 54s - loss: 0.2285 - accuracy: 0.9638
Epoch 100/100
- 53s - loss: 0.2307 - accuracy: 0.9633
```

Fig 4.2: Training data.

4.6 EXPERIMENT ACCURACY, LOSS & OUTPUT

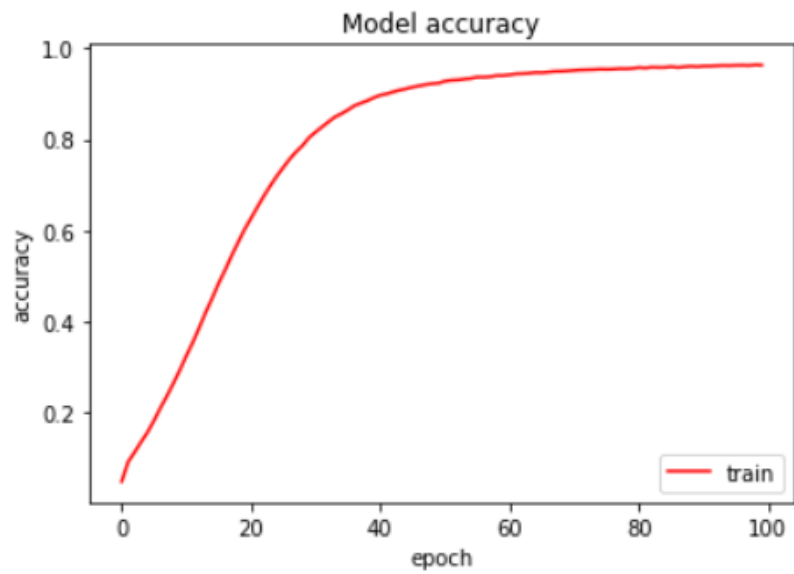


Fig 4.3: Model accuracy.

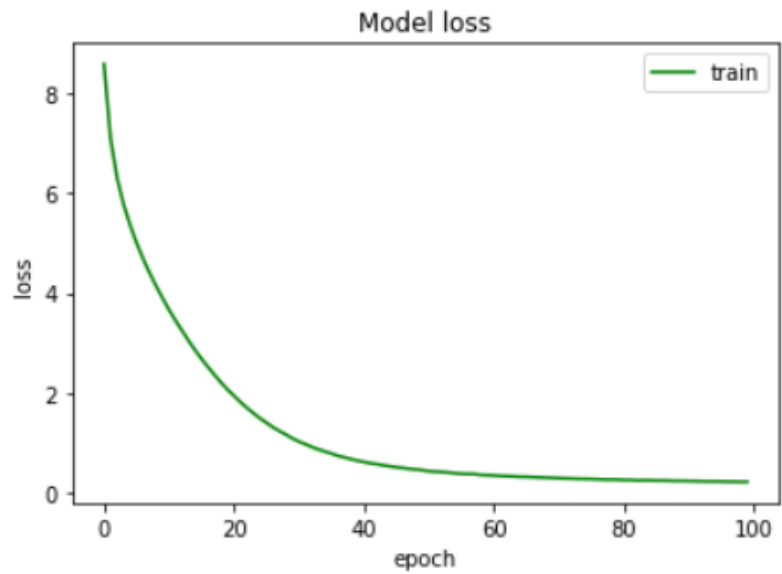


Fig 4.4: Model loss.

```

generate_paragraph("ঝুঁকির মধ্যে দেশ", 200, model, max_sequence_len)
'ঝুঁকির মধ্যে দেশ থেকে সরে আসা ভুল ছিল বলে মনে হচ্ছে না। ভিসা জালিয়াতি এবং গৃহপরিচারিকার বেতন নিয়ে ভুয়া তথ্য দেওয়ার অভিযোগে দেবযানীর
<
generate_paragraph("রাজধানী", 200, model, max_sequence_len)
'রাজধানী ঢাকাকে আবারও সারা দেশ থেকে বিচ্ছিন্ন করল সরকার গত রোববার থেকে টানা দুই দিন সড়কপথ রেলপথ ও নৌপথে ঢাকামুখী যাত্রা ঠেকিয়ে রা
<
generate_paragraph("সুসময় এনে দেবে মানুষ", 200, model, max_sequence_len)
'সুসময় এনে দেবে মানুষ গতকাল মঙ্গলবার বিকেলে ঢাকা বিশ্ববিদ্যালয়ের কার্জন হলের আদলে তৈরি করা হয়েছে বাংলাদেশ প্যাভিলিয়ন এখানে ঠাই পেয়েছে ব
<
generate_paragraph("আজ মওলানা ভাসানী স্টেডিয়ামে জাতীয় স্কুল হকির সাড়ম্বর উদ্বোধন হবে", 200, model, max_sequence_len)
'আজ মওলানা ভাসানী স্টেডিয়ামে জাতীয় স্কুল হকির সাড়ম্বর উদ্বোধন হবে আন্তর্জাতিক হকি ফেডারেশনের সভাপতি লিয়ান্দ্রো নেগের হাত দিয়ে গতকালই ঢাক
<
generate_paragraph("মানুষ", 200, model, max_sequence_len)
'মানুষ নারী নির্যাতন বিষয়ে রাষ্ট্র তথা বিচারব্যবস্থা যত দ্রুত সময়োচিত ব্যবস্থা নেবে তত দ্রুত সুবিচার প্রাপ্তি নিশ্চিত হবে নারী নির্যাতন প্রতিরোধে চলমান
<

```

Fig 4.5: Output.

Input	ঝুঁকির মধ্যে দেশ
Output	ঝুঁকির মধ্যে দেশ থেকে সরে আসা ভুল ছিল বলে মনে হচ্ছে না। ভিসা জালিয়াতি এবং গৃহপরিচারিকার বেতন নিয়ে ভুয়া তথ্য দেওয়ার অভিযোগে দেবযানীর বিরুদ্ধে নিউইয়র্কে মামলা রয়েছে। তিনি কূটনৈতিক দায়মুক্তির সুবিধা কাজে লাগিয়ে ওই অভিযোগ থেকে অব্যাহতি পেতে আইনি লড়াই চালিয়ে যাচ্ছেন। মার্কিন পররাষ্ট্র দপ্তর গত সপ্তাহে বলেছে দেবযানীকে গ্রেপ্তার ও পরবর্তী পদক্ষেপ সম্পর্কে মার্কিন আইনশৃঙ্খলা রক্ষাকারী বাহিনীর ভূমিকা সঠিক ছিল। পররাষ্ট্র দপ্তরের এই সমর্থনের ফলে দেবযানীর বিরুদ্ধে মামলার কার্যক্রম অব্যাহত রাখতে মার্কিন কোঁসুলি প্রীত ভারারার কোনো আইনি বাধা থাকছে না। ফলে দায়মুক্তি সুবিধার আওতায় দেবযানীর বিরুদ্ধে অভিযোগ প্রত্যাহারের সম্ভাবনা নেই। এ ব্যাপারে ভারারা ইতিমধ্যে দেবযানীর বিরুদ্ধে প্রয়োজনীয় আর্টিকল দলিলপত্র আদালতে পেশ করেছেন। এর আগে দেবযানীর আইনজীবী ড্যানিয়েল আর্শাক গত জানুয়ারি তাঁর মক্কেলের বিরুদ্ধে গ্রেপ্তারি পরোয়ানা প্রত্যাহার করে তাঁকে দেশে ফেরত পাঠানোর অনুমতি প্রার্থনা করেন। তিনি গত শুক্রবার বলেন মার্কিন কোঁসুলি আসল ঘটনা এবং আইনের ব্যত্যয় ঘটাচ্ছেন। বিষয়টি নিয়ে আদালতই চূড়ান্ত সিদ্ধান্ত দেবেন। দেবযানীকে নিয়ে বিতর্কের অবসানে মার্কিন সরকারের ভূমিকায় কূটনীতিকদের একটি অংশে দ্বিধাভিত্তি রয়েছে। তবে একটি অংশ মনে করছে মার্কিন কর্তৃপক্ষ দেবযানীর বিরুদ্ধে আইনি লড়াই অব্যাহত রাখার যে সিদ্ধান্ত নিয়েছে তাতে দ্বিপক্ষীয় সম্পর্কের আরও অবনতির আশঙ্কা রয়েছে। ভারতের সাবেক কূটনীতিকদের একটি অংশ মনে করছেন দেবযানীকে নিয়ে দুই দেশের মধ্যে বিরোধ

	কূটনৈতিকভাবেই নিষ্পত্তির সুযোগ ছিল। যুক্তরাষ্ট্রের আইনি অবস্থান কিছুটা ব্যতিক্রমী দৃষ্টান্ত স্থাপন করেছে।
Input	রাজধানী
Output	রাজধানী ঢাকাকে আবারও সারা দেশ থেকে বিচ্ছিন্ন করল সরকার গত রোববার থেকে টানা দুই দিন সড়কপথ রেলপথ ও নৌপথে ঢাকামুখী যাত্রা ঠেকিয়ে রাখার কারণে একধরনের সরকারি অবরোধ অবস্থা তৈরি হয়েছে এর ফলে সারা দেশের যোগাযোগব্যবস্থা প্রায় অচল হয়ে পড়ে এ অবস্থায় আজ থেকে সারা দেশে অনির্দিষ্টকালের সড়কপথ রেলপথ ও নৌপথে অবরোধ কর্মসূচি ডেকেছে বিরোধী জোট বিএনপির নেতৃত্বাধীন দলীয় জোটের নেতা কর্মীরা সমাবেশ করতে চাইলে এবং পুলিশ ও সরকার সমর্থকেরা তা প্রতিহত করায় ঢাকাসহ সারা দেশে সংঘাতের ঘটনা ঘটে সংঘর্ষে নাটোর চাঁপাইনবাবগঞ্জ ও রাজশাহীতে চারজন নিহত হয়েছেন সারা দেশে আহত হয়েছেন জন নিহত ব্যক্তির সবাই বিএনপির নেতা কর্মী রাতে বিএনপির পাঠানো সংবাদ বিজ্ঞপ্তিতে বলা হয় গতকালের সংঘর্ষে তাদের তিনজন নিহত ও জন আহত হয়েছেন গ্রেপ্তার হয়েছেন প্রায় তবে স্থানীয় বিএনপি চারজন নিহত হওয়ার কথা জানিয়েছে ঢাকা মহানগর পুলিশ ডিএমপি জানিয়েছে গত রবি ও সোমবার রাজধানী থেকে জনকে পুলিশ গ্রেপ্তার করেছে এঁদের মধ্যে পাঁচজনকে ভ্রাম্যমাণ আদালত এক বছর করে কারাদণ্ড দেন বিএনপির নেতৃত্বাধীন দল গতকাল ঢাকায় গণতন্ত্র হত্যা দিবস পালনের কর্মসূচি দিয়েছিল সরকারি দল পাল্টা কর্মসূচি ঘোষণা করে গণতন্ত্রের বিজয় দিবস নামে জানুয়ারির একতরফা নির্বাচনের প্রথম বার্ষিকী উপলক্ষে সরকার ও বিরোধী পক্ষ এ মারমুখী অবস্থান নিয়েছে শুক্র ও শনিবারের সাপ্তাহিক ছুটির সঙ্গে রোববার ঈদে মিলাদুন্নবী সা এর ছুটি সরকারি অবরোধ আর বিরোধীদের তা
Input	সুসময় এনে দেবে মানুষ
Output	সুসময় এনে দেবে মানুষ গতকাল মঙ্গলবার বিকেলে ঢাকা বিশ্ববিদ্যালয়ের কার্জন হলের আদলে তৈরি করা হয়েছে বাংলাদেশ প্যাভিলিয়ন এখানে ঠাই পেয়েছে বাংলাদেশের টি প্রকাশনা সংস্থা এর মধ্যে সাতটি সরকারি বইমেলা শুরু হয়েছে গত জানুয়ারি থেকে চলবে ফেব্রুয়ারি পর্যন্ত তম এই আন্তর্জাতিক বইমেলায় বইপ্রেমীদের আকর্ষণের কেন্দ্রে রয়েছে বাংলাদেশ প্যাভিলিয়ন কলকাতার বইপ্রেমীরা লাইন দিয়ে ঢুকছেন কিনছেন তাঁদের পছন্দের বই বাংলা একাডেমির ভাষাসংক্রান্ত নানা ধরনের বই নজরুল ইনস্টিটিউটের বইসহ কম্পিউটার বিষয়ক বিভিন্ন বাংলা বই এবার চাহিদার শীর্ষে প্রখ্যাত লেখক হুমায়ূন আহমেদের উপন্যাসও বেশ বিক্রি হচ্ছে এবারের বইমেলায় বাংলাদেশসহ টি দেশ অংশ নিয়েছে স্টল হয়েছে টি কলকাতা বইমেলায় আগামীকাল ফেব্রুয়ারি পালিত হবে বাংলাদেশ দিবস এ উপলক্ষে এদিন বিকেলে মিলনমেলা ময়দানের এসবিআই মিলনায়তনে আয়োজন করা হয়েছে সভ্যতার জন্য গ্রন্থ গ্রন্থের জন্য সভ্যতা শীর্ষক এক আলোচনা সভার এতে প্রধান অতিথি থাকবেন বাংলাদেশের সংস্কৃতিমন্ত্রী আসাদুজ্জামান নূর মূল প্রবন্ধ পড়বেন বাংলাদেশের শিক্ষাসচিব কবি কামাল আবদুল নাসের চৌধুরী আলোচনা করবেন

	সংস্কৃতিসচিব রঞ্জিত কুমার বিশ্বাস বাংলা একাডেমির মহাপরিচালক শামসুজ্জামান খান নজরুল ইনস্টিটিউটের নির্বাহী পরিচালক কবি রশিদ হায়দার জাতীয় গ্রন্থকেন্দ্রের পরিচালক কবি অসীম সাহা কবি নির্মলেন্দু গুণ বাংলাদেশের জ্ঞান ও সৃজনশীল প্রকাশনা সমিতির সভাপতি ওসমান গনিসহ বিশিষ্টজনেরা আলোচনা শেষে গান গাইবেন বাংলাদেশের শিল্পীরা পরিবেশন করেন শিল্পীরা পরিবেশন করেন সত্যেন সেন শিল্পীগোষ্ঠী ও ক্রান্তি শিল্পীগোষ্ঠী দলীয় নৃত্য পরিবেশন করে নৃত্যশ্রী ধৃতি ও বহিঃশিখা উদ্বোধনী অনুষ্ঠানে বিশেষ অতিথি
Input	মানুষ
Output	মানুষ নারী নির্যাতন বিষয়ে রাষ্ট্র তথা বিচারব্যবস্থা যত দ্রুত সময়োচিত ব্যবস্থা নেবে তত দ্রুত সুবিচার প্রাপ্তি নিশ্চিত হবে নারী নির্যাতন প্রতিরোধে চলমান মামলাগুলোর অবস্থা পর্যালোচনা করে পরবর্তী পদক্ষেপ নিতে হবে দরকার বিদ্যমান আইনের যথাযথ প্রয়োগেরও গতকাল বৃহস্পতিবার বিকেলে বাংলাদেশ মহিলা পরিষদের লিগ্যাল এইড উপপরিষদের উদ্যোগে সুফিয়া কামাল ভবন মিলনায়তনে আয়োজিত এক কর্মশালা ও মতবিনিময় সভায় বক্তারা এ কথা বলেন বাংলাদেশ মহিলা পরিষদ এ কর্মশালার আয়োজন করে বাংলাদেশ মহিলা পরিষদের কেন্দ্রীয় কমিটির সভাপতি আয়শা খানম বলেন নারী নির্যাতন বন্ধের কাজ দুরূহ কাজ মহিলা পরিষদ সাধ্যমতো এ কাজ করে যাচ্ছে নারী নির্যাতন প্রতিরোধে বিদ্যমান আইনের যথাযথ প্রয়োগ নিশ্চিত করতে হবে সংগঠনের ভারপ্রাপ্ত সম্পাদক রাখী দাশ পুরকায়স্থ বলেন আমরা এ দেশে এমন একটি সমাজ ও রাষ্ট্রব্যবস্থা নিশ্চিত করতে চাই যেখানে কোনো শিশু বা নারী নির্যাতনের শিকার হবে না কর্মশালায় সংগঠনের সহসাধারণ সম্পাদক মাসুদা রেহানা লিগ্যাল এইড সম্পাদক সাহানা কবীর প্রমুখ বক্তব্য দেন দুপুরে নারী কর্মশালা কর্মশালা হয়েছে এ পুরস্কার বিতরণ করা হবে উৎসব আয়োজন করে দেওয়া হয়েছে সভায় সভাপতিত্ব করেন প্রধান অতিথি হিসেবে উপস্থিত ছিলেন জেলা আওয়ামী লীগের সভাপতি সাংসদ ওমর ফারুক চৌধুরী সাধারণ সম্পাদক আসাদুজ্জামান আওয়ামী লীগের নেতা নুরুল ইসলাম ইব্রাহিম হোসেন আবদুস সামাদ প্রমুখ সভায় উপজেলার শুভডাঙ্গা আউচপাড়া নরদাশ গোবিন্দপাড়া ও সোনাডাঙ্গা ইউনিয়নের বিভিন্ন পেশার লোকজন উপস্থিত ছিলেন না তাঁর গ্রামের ও

Table 4.3: Text Output According to Input.

4.7 SUMMARY

After completion of all the processes successfully we get our anticipated result with 96.33% accuracy rate and loss with 0.2307.

CHAPTER 5

Summary, Recommendation, Conclusion and Future Work

5.1 SUMMARY OF RESEARCH

The research project is for generating paragraph text. For that we needed Bengali dataset. Datasets are collected from various online sources. After that data is processed to feed into the system. Finally cleaning the data, we feed it into the system to train and the output generates paragraphs.

5.2 RECOMMENDATION

Other models can also be used to understand the text dependent on the dataset. Since, understanding the text dependent on the dataset is important. It might change the whole idea of what has been understood before. Different and various data gives various possibilities of outcome.

5.3 CONCLUSION

In this paper, we used Bangla language and we proposed a paragraph generation technique that can generate a paragraph of a fixed length from a given title. We used Bi-directional LSTM for generating a paragraph using the next word prediction method. Our model gives satisfaction output. As there is some oversight in this model we'll try to overcome this with our future work.

5.4 FUTURE WORK

As our proposed model has some oversight, for example our model gives us a paragraph using next word prediction but it will not be perfect in all cases. We will try to overcome this in future. We will make a model that can be able to predict the next sentence as well as the next word. Then we will predict a sentence from the given title and after that we will take our predicted sentence as input to predict sequential sentences. We believe it will be more perfect than this. We think we could also overcome our problem of fixed length paragraphs.

References

- [1] Md. Sanzidul Islam, Sadia Sultana Sharmin, Sheikh Abujar and Syed Akhter Hossain. "Sequence-to-Sequence Bangla Sentence Generation with LSTM Recurrent Neural Networks." *Procedia Computer Science*, vol. 152, 2019, pp. 51–58.
- [2] Md Ashraful Islam Talukder, Sheikh Abujar, Abu Kaisar Mohammad Masum, Fahad Faisal, Syed Akhter Hossain et al. "Bengali Abstractive Text Summarization Using Sequence to Sequence RNNs." *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, pp. 1–5.
- [3] Sheikh Abujar, Abu Kaisar Mohammad Masum, S. M. Mazharul Hoque Chowdhury, Mahmudul Hasan, Syed Akhter Hossain. "Bengali Text Generation Using Bi-Directional RNN." *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, pp. 1–5.
- [4] P. P. Barmana, A. Boruah, "A RNN based Approach for next word prediction in Assamese Phonetic Transcription," In *Procedia Computer Science*, Volume 143, Pages 117-123, 2018.
- [5] H. Noh, P. H. Seo, B. Han, "Image Question Answering Using Convolutional Neural Network With Dynamic Parameter Prediction," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30-38, 2016.
- [6] M. Sundermeyer, R. Schlüter, H. Ney, "LSTM Neural Networks for Language Modeling," In *13th Annual Conference of the International Speech Communication Association*, pp. 194-197, Portland, OR, USA, September 2012.
- [7] T. Mikolov, W. Yih, G. Zweig, "Linguistic Regularities in Continuous Space Word Representations," In *Proceedings of NAACL-HLT 2013*, pages 746–751, Atlanta, Georgia, 9–14 June 2013.
- [8] Sanzidul Islam, et al. "Sequence-to-sequence Bangla Sentence Generation with LSTM Recurrent Neural Networks." *Procedia Computer Science* 152 (2019): 51-58.

- [9] Amir H. Jadidinejad, Neural Machine Transliteration: Preliminary Results, arXiv:1609.04253v1 [cs.CL] 14 Sep 2016.
- [10] Xingxing Zhang and Mirella Lapata, Chinese poetry generation with recurrent neural networks, In Proceedings of the 2014 Conference on EMNLP, Association for Computational Linguistics, October, 2014.
- [11] Praveen Krishnan, Naveen Sankaran T, Ajeet Kumar Singh, C V Jawahar, Towards a Robust OCR System for Indic Scripts, International Workshop on Document Analysis Systems, Centre for Visual Information Technology, International Institute of Information Technology Hyderabad - 500 032, INDIA, April 2014.
- [12] Robert Ostling and Johannes Bjerva, SU-RUG at the CoNLLSIGMORPHON 2017 shared task: Morphological Inflection with Attentional Sequence-to-Sequence Models, arXiv:1706.03499v1 [cs.CL] 12 Jun 2017.
- [13] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, S Khudanpur, “Extensions of recurrent neural network language model,” In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), July 2011.
- [14] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, Christian Jauvin, “A Neural Probabilistic Language Model,” In The Journal of Machine Learning Research, Volume 3, Pages 1137-1155, February 2003.
- [15] Ramesh Nallapati, Bowen Zhou, et al “Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond”.The SIGNLL Conference on Computational Natural Language Learning (CoNLL), 26 Aug 2016.
- [16] Peter J. Liu et al. “Generating Wikipedia by Summarizing Long Sequences”. International Conference on Learning Representation (ICLR), 2018.
- [17] Lifeng Shang, Zhengdong Lu, Hang Li “Neural Responding Machine for Short-Text Conversation”. Association for Computational Linguistics (ACL 2015)

- [18] Minh-Thang Luong, Ilya Sutskever, Quoc V. Le, Oriol Vinyals, Wojciech Zaremba “Addressing the Rare Word Problem in Neural Machine Translation”. Association for Computational Linguistics (ACL, 2015)
- [19] Ilya Sutskever, James Martens and Geoffrey E. Hinton, Generating text with recurrent neural networks, In Proceedings of the 28th International Conference on Machine Learning (ICML-11), ACM, 2011
- [20] Francois Mairesse, Milica Gasic, Filip Jurccek, Simon Keizer, Blaise Thomson, Kai Yu and Steve Young, Phrase-based statistical language generation using graphical models and active learning, In Proceedings of the 48th ACL, ACL 10, 2010.
- [21] Francois Mairesse and Steve Young, Stochastic language generation in dialogue using factored language models, Computer Linguistics, 2014
- [22] Tomas Mikolov, Martin Karafit, Lukas Burget, JanC ernocky, and Sanjeev Khudanpur, Recurrent neural network based language model, In Proceedings on InterSpeech, 2010

Bengali Paragraph Generation Using Bi-directional LSTM

ORIGINALITY REPORT

12%

SIMILARITY INDEX

3%

INTERNET SOURCES

10%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Sheikh Abujar, Abu Kaisar Mohammad Masum, S. M. Mazharul Hoque Chowdhury, Mahmudul Hasan, Syed Akhter Hossain. "Bengali Text generation Using Bi-directional RNN", 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019
Publication 7%
- 2 Md. Sanzidul Islam, Sadia Sultana Sharmin Mousumi, Sheikh Abujar, Syed Akhter Hossain. "Sequence-to-sequence Bangla Sentence Generation with LSTM Recurrent Neural Networks", Procedia Computer Science, 2019
Publication 2%
- 3 dspace.daffodilvarsity.edu.bd:8080
Internet Source 1%
- 4 research.google.com
Internet Source 1%
- 5 Md Ashraful Islam Talukder, Sheikh Abujar, Abu Kaisar Mohammad Masum, Fahad Faisal, Syed <1%

Akhter Hossain. "Bengali abstractive text summarization using sequence to sequence RNNs", 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019

Publication

6 Abu Kaisar Mohammad Masum, Sheikh Abujar, Raja Tariqul Hasan Tusher, Fahad Faisal, Syed Akhter Hossain. "Sentence Similarity Measurement for Bengali Abstractive Text Summarization", 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019

Publication

7 "Innovations in Computer Science and Engineering", Springer Science and Business Media LLC, 2020

Publication

8 "Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2018

Publication

9 Wilhelm Kirchgassner, Oliver Wallscheid, Joachim Bocker. "Deep Residual Convolutional and Recurrent Neural Networks for Temperature Estimation in Permanent Magnet Synchronous Motors", 2019 IEEE International Electric

Machines & Drives Conference (IEMDC), 2019

Publication

10 "Machine Learning and Information Processing", Springer Science and Business Media LLC, 2020 **<1%**
Publication

11 arxiv.org **<1%**
Internet Source

12 Ngoc Khuong Nguyen, Anh-Cuong Le, Hong Thai Pham. "Chapter 22 Deep Bi-directional Long Short-Term Memory Neural Networks for Sentiment Analysis of Social Data", Springer Science and Business Media LLC, 2016 **<1%**
Publication

13 www.aclweb.org **<1%**
Internet Source

Exclude quotes On

Exclude matches < 3 words

Exclude bibliography On