

**A DEEP LEARNING APPROACH FOR GOVERNMENT AUTOMATED  
BENGALI NOTE WRITING SERVICE**

**BY**

**Md. Muhaiminul Islam  
ID: 163-15-8473**

**Mohiyminul Islam  
ID: 172-15-10187  
AND**

**Md. Abu Yousuf  
ID: 163-15-8480**

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

Supervised By

**Dr. Fizar Ahmed**  
Assistant professor  
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**

**MAY 2021**

**APPROVAL**

This Project titled “**A deep learning approach for government automated Bengali note writing service**”, submitted by **\*Md. Muhaiminul Islam\*** **\*Mohiyminul Islam\*** and **\*Md. Abu Yousuf\*** 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 **\*31-05-2021\***.

**BOARD OF EXAMINERS**

**Chairman**



---

**Dr. Touhid Bhuiyan**

**Professor and Head**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



---

**Internal Examiner**

**Nazmun Nessa Moon**

**Assistant Professor**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



**Internal Examiner**

---

**Aniruddha Rakshit**

**Senior Lecturer**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



**External Examiner**

---

**Dr. Mohammad Shorif Uddin**

**Professor**

Department of Computer Science and Engineering

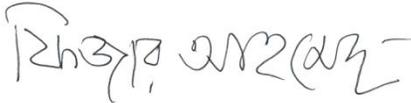
Jahangirnagar University

## DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Fizar Ahmed, Assistant professor, Department of CSE** 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:

**Dr. Fizar Ahmed**



---

Assistant professor  
Department of CSE  
Daffodil International University

### Co-Supervised by:

**Md. Tarek Habib**

---

Assistant professor  
Department of CSE  
Daffodil International University

### Submitted by:

**Md. Muhaiminul Islam**

ID: 163-15-8473  
Department of CSE  
Daffodil International University

**Mohiyminul Islam**

ID: 172-15-10187

Department of CSE

Daffodil International University

**Md. Abu Yousuf**

ID: 163-15-8480

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 successfully.

We really grateful and wish our profound our indebtedness to **Dr. Fizar Ahmed, Assistant professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Natural Language Processing*” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Prof. Dr. Touhid Bhuiyan, Professor and Head**, Department of CSE, for his kind help to finish our project and also to other faculty member 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 discuss while completing the course work.

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

## **ABSTRACT**

The world is undergoing various changes despite the revolution of modern technologies. With the help of automation tools, human tasks are been made very simple and automated. People are adopting Artificial intelligence in healthcare, retail and E-commerce, food and tech, travel, banking and financial services, entertainment, and various sectors in the developed countries. But in the underdeveloped or developing countries still, now there are some issues to solve. In this world, there are approximately 228 million people who communicate in the Bengali language. And in Bangladesh Bengali is also used as an official language. Govt. or non-govt. offices use Bengali to perform their official tasks. They have to write down their official tasks, letters, or maybe important notes on their own. As there is no automation tool available regarding Bengali note writing, the office staffs have to spend their valuable time by writing down the large notes or documents. This particular thing not only hampers the time but also is a bad usage of valuable human resources. To bridge this gap we have proposed a model that will be able to generate large notes by typing hints of a couple of words. We have used Bag of words and three-layer Artificial neural network to resolve this issue. The Bengali data was processed in a special way so that our model can perform well with about 1431 text documents. After running a couple of experiments our model was able to generate large notes from short notes with 91% accuracy.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	ii-iii
Declaration	iv-v
Acknowledgements	v-vi
Abstract	vi
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-3</b>
<b>CHAPTER 2: LITERATURE REVIEW</b>	<b>3-4</b>
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>4-16</b>
3.1 Problem Assertion	6
3.2 Data Collection	7-10
3.3 Pre-analysis of Data	10-11
3.4 Data Pre-processing	11
3.5 Bag of Words Model	12-13
3.6 Algorithms used for the Experiment	13-16

3.6.1 Decision Tree	14
3.6.2 Random Forest	14
3.6.3 Multilayer Artificial Neural Network	14-16
<b>CHAPTER 4: EXPERIMENTAL RESULTS</b>	<b>17-21</b>
<b>CHAPTER 5: CONCLUSION AND FUTURE WORK</b>	<b>22</b>
<b>CHAPTER 5: REFERENCES</b>	<b>22-24</b>

## LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1: Workflow for the proposed framework	5
Figure 1.2: Demo application	6
Figure 1.3: Number of words in data pattern	11
Figure 1.4: bag of words model sample 1	12
Figure 1.5: bag of words model sample 2	12
Figure 1.6: bag of words model sample 3	13
Figure 1.7: bag of words model sample 4	13
Figure 1.8: bag of words model sample 5	13
Figure 1.9: Softmax equation	15
Figure 2.0: Architecture of proposed model 1	16
Figure 2.1: Architecture of proposed model 2	16
Figure 2.2 : Analysis of results	18

## **LIST OF TABLES**

<b>TABLES</b>	<b>PAGE NO</b>
Table 1.1: Demo of the dataset	7-10
Table 1.2: Results of hyper-parameter tuning	17
Table 1.3: Comparison with other models	17
Table 1.4: Input and output sample 1	18-19
Table 1.5: Input and output sample 2	19
Table 1.6: Input and output sample 3	19-20
Table 1.7: Input and output sample 4	20-21
Table 1.8: Input and output sample 5	21

## **CHAPTER 1**

### **Introduction**

In the era of modern technology Artificial intelligence plays a very important role. Scientists introduced AI to get rid of complex problems which human cannot solve within time. From the 90's various researchers have gone through several studies to adopt the AI in our practical life. This single technology not just only solves complicated

problems but also saves a lot of time for us. Nowadays various important and crucial works are given to AI and it is performing very well according to expectation.

Scientists are in a collaboration with computer vision and Natural Language Processing which are two of the famous branches of AI. However Solving the NLP tasks were very much difficult for AI as there are approximately 6500 different languages in the current world. So it is very much difficult to train the AI with such diverse data. But different region's researchers are trying their best to enrich this NLP field with their own language. But still it was a dark era of NLP till 1950.

From 1950 scientists brought a new revolution to NLP. There were lots of ups and downs in research studies regarding Natural Language Processing. Before the introduction of deep learning techniques working with text or sequence was very much difficult as it was not showing human likely results with Machine learning algorithms. The concept of Neural networks was 1<sup>st</sup> introduced by Walter Pitts and warren Mcculloch in 1943 [1]. Until 1969 this neural network was one of the most major areas of research in the field of computer science and neuroscience. But still, this method failed in many steps to solve different problems. Mainly it brought attention among the top scholars when the backpropagation is proposed by Werbos [2]. After the new revolution of Deep learning the field of Natural Language Processing also got the highest peak. Researchers were creating new techniques to bridge the gap between Machine and human language. In 1997 LSTM network was proposed by Hochreiter and Schmidhuber to reduce the sequence gap and to remember the longer sequence [3].

After evolving several times the computer started to work with the human language quite efficiently. Chatbots, text-generator, text to images, text summarizer many applications started their journey to contribute to humans. NLP is not limited to just English language. Researchers continuously working hard for contributing to their native language. In Bangladesh, Bengali Natural Language Processing (BNLP) is also one of the most major sectors of the AI research field nowadays. Though there is a lack of Language resources of Bengali but researchers trying die-hard to solve these issues. In this paper, the problem statement is quite similar to chatbots. Instead of giving a response, we wanted to produce

documents as texts. Users will just type a simple couple of words and our algorithm will try to classify it into appropriate notes and will return the best match among the documents. So large size documents will be generated with the matter of a very small portion of the time. Notes were written in Bengali. We processed it manually with the help of Python 3.7.0. The bag of words model was used for processing the data and to make readable for the algorithms. At first machine learning algorithms were used to identify the patterns of the texts. But for having a larger dimension of texts, the algorithms could not perform well. That's why we have tried to build our own neural network-based model with various parameters to increase the accuracy. We have successfully done that and came up with a better accuracy score which is 91%. The error rate is very low for our model. We trained the model with a different combination of Hyper-parameters to gain maximum accuracy.

## **CHAPTER 2**

### **Literature review**

From the '90s researchers made various conversational tools like Chatbots, Text-generator. There have been lots of researches regarding sequence to sequence NLP tasks. The initial idea of Chabot was first introduced by Weizenbaum, J. et al[4]. The discussion was about an intelligent Chatbot named ELIZA which is nothing but an advanced computer program also known as the Doctor [5]. By using this Chatbot everyone started to think the computer can really communicate with humans and can hear their problems. ALICE a heuristic pattern matching chatbot was also brought to attention [6]. Jabberwacky a context-based chatbot that studies language and interacts[7]. It was a relational agent that mainly trained on human conversations and comments. A relational agent was introduced by Bickmore et al.[8] to establish social bonds[9]. "Rea" a popular agent shows non-verbal behavior which was introduced by Cassell, J et al [10]. In the year 2020, there was a Machine conversation model that was created by using a Deep neural network [11] authored by G Krishna, Rasool, Gaurav Hajela.

Besides the above works, Bengali researchers also contributed to this field. A chatbot for Bangla education named Doly was introduced in 2019 [12]. They have built their own Bengali corpus of texts and responses. Using two logical adapters they have processed the input texts and generated output. A Bengali informative chatbot based on TF-IDF model and cosine similarity to find out the accurate answers[13]. Bengali anaphora resolution presented by using Hobb's algorithm. They claimed to achieve 97.22% accuracy. Recently in 2020, a team of young researchers introduced a Bengali question answering system which was mainly based on a Bengali general knowledge dataset [14]. It was a Seq2Seq model based on encoder and decoder. Their system performed with 89% accuracy on the validation data. Corpus was also built by them. Besides these studies, nowadays Bengali researchers are very initiative to develop Bengali AI resources. Young Gang Cao et al. introduced a model named AskHERMES which was answering clinical questions and there was a dataset of complex questionnaires [17]. Somnath Banerjee et al. proposed a Bengali QA system by using question classification system [18]. L. HIRSCHMAN and R. GAIZAUSKAS explained about the question answering system and put a review on different approaches and their pros and cons [19]. Chandra Obula Reddy and Dr. K. Madhavi also showed a survey on different question answering techniques and how they performed [20].

All of these studies contributed to generate text to text results. And Bengali researchers are also coming forward to enrich Bengali models

## **CHAPTER 3**

### **Research methodology**

This particular research study has gone through a special pipeline to generate fruitful output. From collecting and analyzing the data followed by some practical experiments led this study to success.

For this research study, we have used python 3.7.0 and Tensorflow 2.2.4. And for performing hyper-parameter tuning we have used Keras Tuner which is a tool for selecting Parameters and layers for the neural network.

To analyze the data and running the experiments we needed GPU support which we have gotten by Google Colaboratory. For developing a primary application we also needed to use Flask (Python backend framework) and for the frontend design we have used JavaScript's library JQuery. The approach started from data collection till the final experiment until we got our desired output from the research. The entire methodology is briefly described in Figure 1.1.

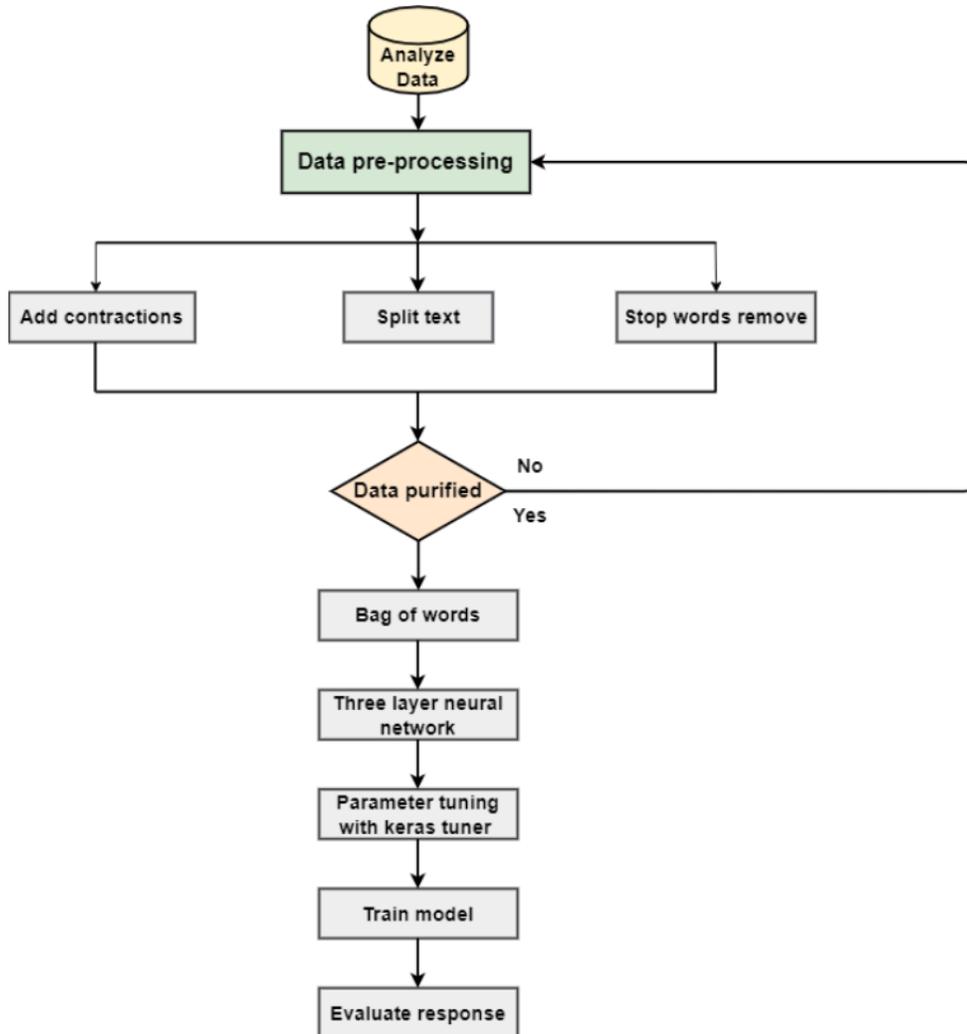


Figure 1.1: Workflow for the proposed framework

### 3.1 Problem assertion

We had to analyze the entire situation before implementing the model. We have noted down the pros. And cons. Our objective was to create a useful model that will be able to generate Large documents by short texts. If we consider the pattern of the text as T and documents as D we can write the concept as  $T < D$ .

There was a language barrier. Working with the Bengali language where there is a lack of experimental tools was the main challenging task. For describing the task we developed a demo application which was made by Flask. The front-end designed with JQuery and backend was handled by Flask. We have trained the bot with our own corpus. Also kept the option for Edit, Save, and printing the documents. The demo was described briefly in Figure 1.2.

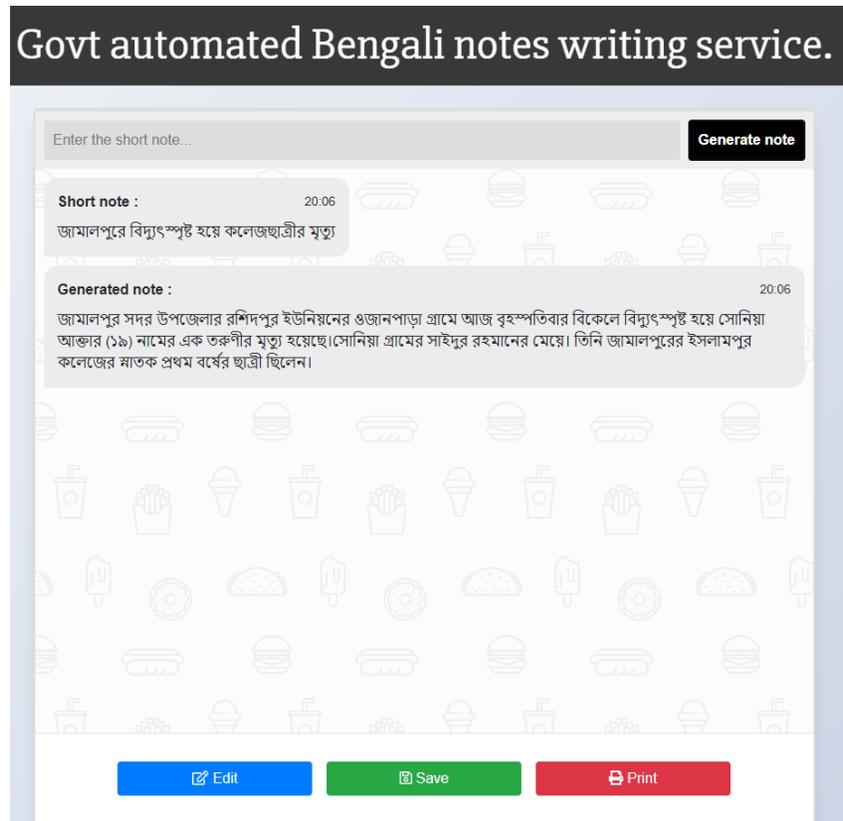


Figure 1.2: Demo application

### 3.2 Data collection

We needed to build our own Bengali corpus as there was no dataset available regarding our problem. We have collected a total of 1441 documents. And we had to reassemble it in such a pattern so the algorithm can easily track it. The source of the data was online Bengali newspapers such as Prothom Alo, Jugantor, Kaler kontho, etc. We kept it in JSON format. In our dataset, there were three fields of each data. “Tag”, “Patterns” and “Responses”. Tags are unique for each of the data. Responses are used as large notes. Patterns for short texts that mean the input. And finally, the “tag” by which the algorithm will find out that in which category does this pattern match and it will generate the corresponding response of that tag. The demo of the dataset is given in table 1.1.

Data 1 :	<pre>{ "tag": "নিহত",   "patterns": ["শাপলা চত্বরে বাসের ধাক্কায় নিহত", "চত্বরে বাসের ধাক্কায় নিহত", "শাপলা চত্বরে বাসের ধাক্কায়"],   "responses": ["রাজধানীর মতিঝিলের শাপলা চত্বর এলাকায় আজ বৃহস্পতিবার সন্ধ্যায় যাত্রীবাহী বাসের ধাক্কায় রঞ্জন দেবনাথ (৩৯) নামের এক ব্যক্তি নিহত হয়েছেন। নিহত রঞ্জ ন দেবনাথের স্ত্রী বীণা বিশ্বাস ঢাকা মেডিকেল কলেজ হাসপাতালের স্টাফ নার্স। তিনি জা নান, রঞ্জন দেবনাথ আফতাব গ্রুপে চাকরি করেন। " }</pre>
Data 2 :	<pre>{ "tag": "বিএনপি",   "patterns": ["অনুমতি না পেলেও সমাবেশ করবে বিএনপি", "অনুমতি না পেলেও সমাবে শ করবে", "সমাবেশ করবে বিএনপি"],   "responses": ["আগামী ৫ জানুয়ারি অনুমতি না পেলেও রাজধানীতে বিএনপি সমাবেশ ক রবে বলে জানিয়েছেন দলের যুগ্ম মহাসচিব রুহুল কবির রিজভী। আজ বৃহস্পতিবার সন্ধ্যা য় জানতে চাইলে প্রথম আলোকে রিজভী জানান, 'আমরা কর্মসূচি ঘোষণা করেছি, শান্তিপূ র্ণভাবে এ কর্মসূচি পালন করব।' }</pre>

Data 3 :	{ "tag": "আগুন", "patterns": ["ইসলামবাগে প্লাস্টিক কারখানায় আগুন", "প্লাস্টিক কারখানায় আগুন", "ইসলামবাগে প্লাস্টিক কারখানায়"], "responses": ["রাজধানীর পুরান ঢাকার ইসলামবাগে একটি প্লাস্টিক পণ্য তৈরির কারখানায় আগুন লেগেছে। আজ বৃহস্পতিবার রাত ৯টা ৫৫ মিনিটে বাগানবাড়ি এলাকায় এই অগ্নিকাণ্ডের ঘটনা ঘটে। ফায়ার সার্ভিসের নিয়ন্ত্রণকক্ষে দায়িত্বরত কর্মকর্তা ফরহাদুজ্জামান প্রথম আলোকে জানান, আগুন নেভাতে ফায়ার সার্ভিসের পাঁচটি ইউনিট কাজ করছে। "] }
Data 4 :	{ "tag": "খুন", "patterns": ["ফকিরাপুলে আবাসিক হোটেলে খুন", "আবাসিক হোটেলে খুন", "ফকিরাপুলে আবাসিক হোটেলে" ], "responses": ["রাজধানীর ফকিরাপুলে আজ বৃহস্পতিবার সন্ধ্যা সাতটার দিকে একটি আবাসিক হোটেলে দুর্বৃত্তদের গুলিতে আবদুল হামিদ (৪৪) নামের এক ব্যক্তি খুন হয়েছেন। হামিদের বাড়ি চট্টগ্রাম জেলার সন্দ্বীপোপুলিশ ও ঢাকা মেডিকেল কলেজ হাসপাতাল সূত্রে জানা যায়, আবদুল হামিদ সপ্তাহ খানেক আগে মতিঝিল থানাধীন ফকিরাপুল এলাকার হোটেল মতিঝিলের ২০১ নম্বর কক্ষ ভাড়া নেন।" ] }
Data 5 :	{ "tag": "বিলবোর্ড", "patterns": ["শনিবারের মধ্যে অবৈধ বিলবোর্ড সরাতে পুলিশের নির্দেশ", "অবৈধ বিলবোর্ড সরাতে পুলিশের নির্দেশ", "শনিবারের মধ্যে অবৈধ বিলবোর্ড" ], "responses": ["চট্টগ্রাম নগরে স্থাপিত অবৈধ সব বিলবোর্ড আগামী শনিবারের মধ্যে অপসারণ করতে বলা হয়েছে। অন্যথায় পরের দিন থেকে বিলবোর্ড স্থাপনকারী মালিকদের গ্রেপ্তার করে আইনি ব্যবস্থা নেওয়া হবে বলে মহানগর পুলিশের পাঠানো এক বিজ্ঞপ্তিতে জানানো হয়েছে।" ] }
Data 6 :	{ "tag": "নির্বাচন", "patterns": ["নির্বাচন কমিশন ছাড়া কারও সঙ্গে সংলাপ নয়: নাসিম", "নির্বাচন কমিশন ছাড়া কারও সঙ্গে সংলাপ নয়", "কারও সঙ্গে সংলাপ নয়: নাসিম" ], "responses": ["আওয়ামী লীগের সভাপতিমণ্ডলীর সদস্য ও স্বাস্থ্যমন্ত্রী মোহাম্মদ নাসিম বিএনপির চেয়ারপারসন খালেদা জিয়ার সংলাপের আহ্বান প্রত্যাখ্যান করে বলেছেন, বিএনপির সঙ্গে সংলাপের প্রশ্নই ওঠে না। নির্বাচন কমিশন ছাড়া কারও সঙ্গে কোনো সংলাপ হবে না। আজ বৃহস্পতিবার বিকেলে রাজধানীর বঙ্গবন্ধু অ্যাভিনিউর ঢাকা মহানগর আওয়ামী লীগের কার্যালয়ে এক বর্ধিত সভায় প্রধান অতিথির বক্তব্যে নাসিম এ কথা বলেন।" ] }

Data 7 :	{ "tag": "বিদ্যুৎস্পৃষ্ট", "patterns": ["জামালপুরে বিদ্যুৎস্পৃষ্ট হয়ে কলেজছাত্রীর মৃত্যু", "কলেজছাত্রীর মৃত্যু", "জামালপুরে বিদ্যুৎস্পৃষ্ট হয়ে"], "responses": ["জামালপুর সদর উপজেলার রশিদপুর ইউনিয়নের ওজানপাড়া গ্রামে আজ বৃহস্পতিবার বিকেলে বিদ্যুৎস্পৃষ্ট হয়ে সোনিয়া আক্তার (১৯) নামের এক তরুণীর মৃত্যু হয়েছে। সোনিয়া গ্রামের সাইদুর রহমানের মেয়ে। তিনি জামালপুরের ইসলামপুর কলেজের স্নাতক প্রথম বর্ষের ছাত্রী ছিলেন।"] }
Data 8 :	{ "tag": "মহাসড়কের", "patterns": ["ঢাকা-চট্টগ্রাম ও ঢাকা-ময়মনসিংহ মহাসড়কের কাজ জুনে শেষ হবে", "ঢাকা-চট্টগ্রাম ও ঢাকা-ময়মনসিংহ মহাসড়কের কাজ", "মহাসড়কের কাজ জুনে শেষ হবে" ], "responses": ["ঢাকা-চট্টগ্রাম ও ঢাকা-ময়মনসিংহ মহাসড়ক চার লেনে উন্নীত করার কাজ আগামী জুনের মধ্যেই শেষ হবে বলে জানিয়েছেন প্রধানমন্ত্রী শেখ হাসিনা। তিনি বলেন, অবকাঠামো নিয়ে কিছু সমস্যা আছে। সেগুলো সমাধানের চেষ্টা চলছে। রেল, সড়ক ও নৌপথকে সমান গুরুত্ব দিয়ে সারা দেশকে একটি নেটওয়ার্কের মধ্যে আনার উন্নয়ন পরিকল্পনা গ্রহণ করা হয়েছে।"] }
Data 9 :	{ "tag": "নবজাতক", "patterns": ["চুরি হওয়া নবজাতক ফিরে গেল মায়ের কোলে", "ফিরে গেল মায়ের কোলে", "চুরি হওয়া নবজাতক"], "responses": ["রাজশাহী মেডিকেল কলেজ হাসপাতাল থেকে গত সোমবার চুরি হওয়া শিশুটিকে মায়ের কোলে ফিরিয়ে দিয়েছে মহানগর পুলিশ। আজ বৃহস্পতিবার সন্ধ্যায় রাজশাহীর পবা উপজেলা থেকে উদ্ধার করা হয় শিশুটিকে। পরে সংবাদ সম্মেলনে করে মহানগর পুলিশের কমিশনার মোহাম্মদ শামসুদ্দিন শিশুটিকে তার বাবা-মায়ের হাতে তুলে দেন। সন্তানকে ফিরে পেয়ে মা রুবিনা খাতুন আনন্দে কেঁদে দেন। বাবার চোখেও ছিল খুশির অশ্রু।"] }
Data 10 :	{ "tag": "কোয়ারেন্টাইনে", "patterns": ["সাদমান হোম কোয়ারেন্টাইনে চলে যান", "কোয়ারেন্টাইনে চলে যান", "সাদমান হোম কোয়ারেন্টাইনে"], "responses": ["বাঁ হাতের কজ্জিতে সফল অস্ত্রোপচার শেষে ১৭ মার্চ অস্ট্রেলিয়া থেকে দেশে ফেরেন সাদমান ইসলাম। ফিরেই বাংলাদেশ ক্রিকেট বোর্ডের (বিসিবি) নির্দেশনা মেনে হোম কোয়ারেন্টাইনে চলে যান তিনি।"] }

Data 11 :	<pre> {"tag": "পরিবহনচালক", "patterns": ["চৌগাছায় চাচাতো ভাইদের হামলায় আহত পরিবহনচালক নিহত", "আহত পরিবহনচালক নিহত", "চাচাতো ভাইদের হামলায়"], "responses": ["যশোরের চৌগাছায় জমি নিয়ে দ্বন্দ্ব চাচাতো ভাইদের হামলায় আহত মুক্তার হোসেন (৫৫) নামে এক পরিবহনচালক মারা গেছেন।শুক্রবার ভোরে ঢাকা মেডিকেল কলেজ হাসপাতালে চিকিৎসাধীন তার মৃত্যু হয়নিহত মুক্তার হোসেন উপজেলার মসিয়ুরনগর গ্রামের জাহাবক্সের ছেলে। তিনি চৌগাছা-ঢাকা রুটে চলাচলকারী রয়েল পরিবহনের চালক ছিলেন।"] } </pre>
-----------	--

Table 1.1: Demo of the dataset

### 3.3 Pre analysis of data

The pre analysis of dataset is always an important step when you are working with a data driven technique. In our case the data was the most vital thing to care about. Without proper understanding the dataset the experiment could not be started. Also when there is need of pre-processing the data one has to completely analyze the patterns of the data otherwise there is very good chance that the results will face some bias issues.

For cleaning and pre-processing the data, we need to understand the characteristics of the data at first. Without proper analysis, we cannot clean the data because we don't know which part is important and which part is not. That's why we have analyzed the input data properly before processing it. Without proper analysis, we cannot clean the data because we don't know which part is important and which part is not. After analyzing properly we have found that the maximum patterns were in between two to eight words. Fewer of them are greater than eight. There were in total of 383 "tags" and 1831 unique words in our dataset considering the "patterns". Figure 1.3 describes well.

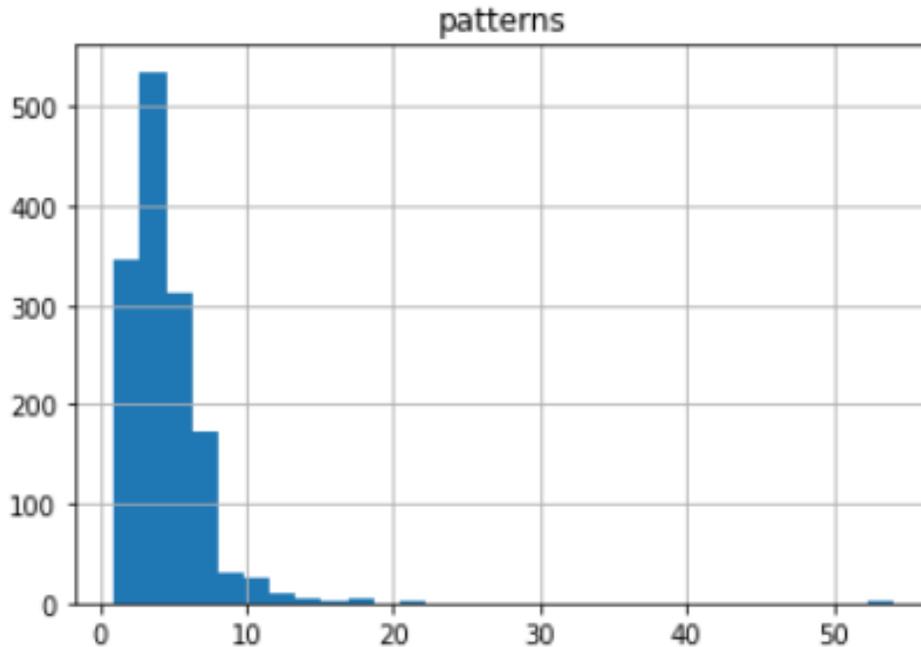


Figure 1.3: Number of words in data pattern

### 3.4 Data pre-processing

If we have to generate maximum output from an ML algorithm or multilayer neural network, the data needs to be pre-processed in a certain way. But important text processing libraries such as NLTK or Spacy do not have resources for the Bengali language. That's why we have undergone some steps. We have pre-processed our data using python programming manually. Contractions have been added to the patterns. Unwanted characters and numbers were removed from the texts as they do not perform any important role in this situation. We have collected some Bengali stop-words and remove them from our dataset as they are not necessary. Ultimately we have shortened the dimension of the input data as much as we can because with higher dimension data algorithm might not perform well.

### 3.5 Bag of words model

Machines just cannot understand human language. They only get the numerical values. That's why need to transform the text in such format so that machine can easily understand. There are some popular text transformation methods. Bag of words, TF-IDF, Word embedding, etc. In these three methods word embedding is most useful and updated. But for Bengali, there is no standard pre-trained word embedding file available. So we have decided to go with the Bag of words model. The model is nothing but a 0 and 1 representation of the texts. It represents the occurrence of words in a document. Each of the words can be expressed in such a sequence. We had to declare a vocabulary list from the corpus. From that, we have created such a sequence of sentences where each of the sentences is a 2D vector of 1831 dimensions. We have applied the Bag of words technique on both input and output data. The “patterns” have been transformed in 1831 dimensional sequence and each of the “tags” has been transformed in 383-dimensional sequence. Data transformation is clearly shown in Figure 1.4-1.8.

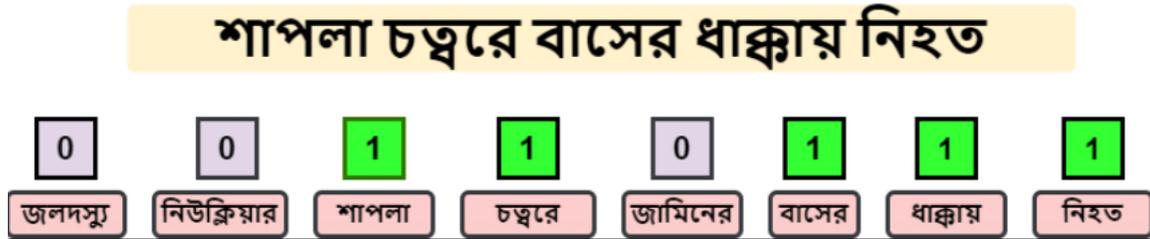


Figure 1.4: Bag of words model sample 1

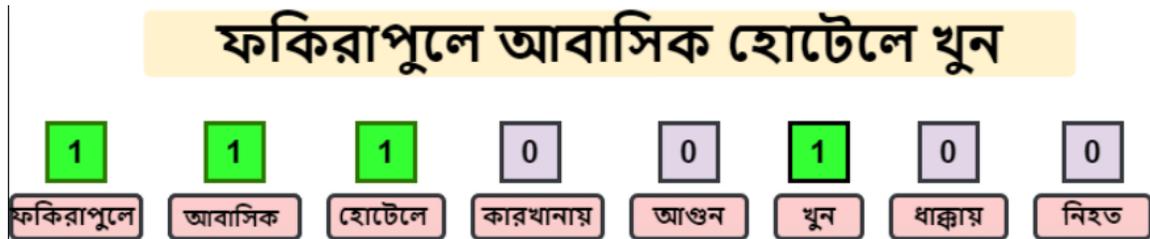


Figure 1.5: Bag of words model sample 2

## কারও সঙ্গে সংলাপ নয়



Figure 1.6: Bag of words model sample 3

## সাদমান হোম কোয়ারেন্টাইনে চলে যান

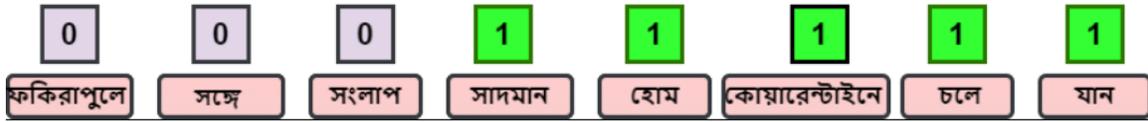


Figure 1.7: Bag of words model sample 4

## ইসলামবাগে প্লাস্টিক কারখানায় আগুন



Figure 1.8: Bag of words model sample 5

### 3.6 Algorithms used for the experiment

Different types of experiments were performed for gaining maximum accuracy. At first, we have tried with simple machine learning concept with a Tree-based algorithm and the Ensemble technique. The decision tree and Random forest algorithm was the first experiment. But having a larger dimension of data they couldn't perform very well. As a consequence, we have gone for a deep learning approach with a three-layer Artificial Neural Network. All of the parameters were selected based on the experience.

### **3.6.1 Decision tree (DT)**

This is one of the most common algorithms used nowadays. This tree-based algorithm predicts its output by maintaining some decision rules. It creates several nodes by considering the features of the dataset. For each of the features, it calculates the information gain and works on it. The node which gains the maximum amount of information and more related to the output is kept as the root node. However, we used this algorithm along with some selected hyper-parameters. We kept criterion as “gini” and `min_samples_split = 2`.

### **3.6.2 Random forest (RF)**

This is basically an ensemble technique which is nothing but a combination of decision trees. Each of the individual tree is created considering feature randomness and bagging. Takes decision from each of the created trees and finally generates output from the majority votes which makes the output more accurate. We kept number of estimator as 90 and `min_samples_leaf` as 1.

### **3.6.3 Multilayer Artificial Neural Network**

Machine learning-based models could not perform well for having such a dimensional(1831) dataset. That's why we have intended to use the deep learning technique. We had a couple of options to choose from over here. We could use a simple ANN or LSTM model for our problem. Generally for text data LSTM performs well because it can store and memorize longer sequences and process the text considering that. But in this case, the pattern of the input text was very short. Each of the input data was mostly two to eight words in length. So for having a smaller sequence we did not introduce LSTM over here, whereas LSTM increases computational expenses. ANN was more than enough to solve our issue. So we have built a multilayer neural network where

there were 328 neurons in the first hidden layer and 168 neurons in the second hidden layer. The total trainable parameter was 720,895. For activating the corresponding neurons we have to apply any kind of activation function. In these circumstances, we have applied "Relu". This function throws its output according to the given input [15]. We can easily express this with equation 1

$$R(z) = \max(0, x) \quad (1)$$

Here  $x$  is the input. The function will return the maximum between 0 and  $x$ . If the  $x$  is negative it will return 0 and if the input is greater than 1 it will directly return the input.

As there are 383 output "tags" in our model so we have used the softmax activation function in the output layer. Softmax will return the probability of each of the documents according to the pattern match. Figure 1.9 briefly describes the softmax [16].

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, \dots, K$$

Figure 1.9: Softmax equation

Each of the neurons in the hidden layers process output in a certain way. Weights get multiplied with the inputs adding a bias with it. We can express the formula with equation 2.

$$y = [x1.w1 + x2.w2 + x3.w3] + b \quad (2)$$

Where  $x1, x2, x3$  are the inputs and  $w1, w2$  and  $w3$  are the weights initialized. Bias is represented by  $b$ . After feeding the neurons with weights in a forward way we get our  $\hat{y}$ . According to  $\hat{y}$  (prediction) and the actual output  $y$ , the mean squared error ( $y - \hat{y}$ ) is calculated. Considering this, the weights start to get updated in a backward way so the error gets minimized as much as possible. During the backpropagation, the weights update in a certain system so the error decreases every time by each update. We have tuned the model with Keras Tuner which is an outstanding tool of Hyper-parameter

tuning for the neural network. According to the tuning result, we have initialized the parameters. Our model's diagram is given in Figure 2.0 and 2.1.

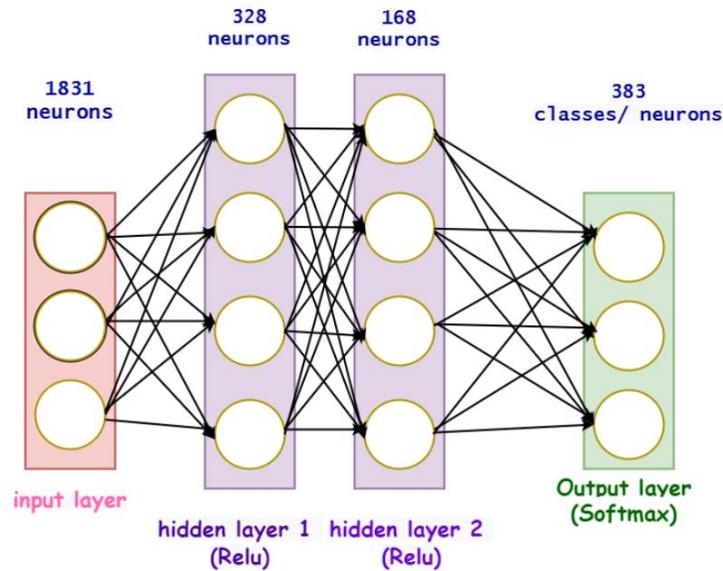


Figure 2.0: Architecture of our proposed model

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 328)	600896
dense_1 (Dense)	(None, 168)	55272
dense_2 (Dense)	(None, 383)	64727
Total params: 720,895		
Trainable params: 720,895		
Non-trainable params: 0		

Figure 2.1: Architecture of our proposed model

## CHAPTER 4

### Experimental results

For building such an accurate neural network we had to go through various experiments. During the training, the right parameters should be selected for better results. We tune the parameters of the neural net such as the number of hidden layers, number of neurons in each layer, number of epochs for training, learning rate, and the perfect combination of activation functions. These important things were selected based on the experiment of Keras tuner and validated by accuracy. We performed a search for around 5.5 hours and finally got the best combinations which were Number of layers =2, Number of neurons in the 1<sup>st</sup> hidden layer=328, Number of neurons in the 2<sup>nd</sup> hidden layer=168, Learning rate=0.01, and Epochs= 150. Table 1.2 describes it in brief.

Rank	Number of hidden layers	Neurons in 1 <sup>st</sup> hidden layer	Neurons in 2 <sup>nd</sup> hidden layer	Neurons in 3 <sup>rd</sup> hidden layer	Learning rate	Epochs/Tuner
<b>01</b>	2	328	168	-----	0.01	150
<b>02</b>	2	264	264	-----	0.001	150
<b>03</b>	2	328	168	-----	0.01	50
<b>04</b>	2	264	264	-----	0.001	50
<b>05</b>	3	456	392	296	0.001	150

Table 1.2: Results of hyper-parameter tuning

After selecting the best parameters our ANN model beat Machine learning models quite significantly. A comparison between the experimental models is shown in the table 1.3.

Rank	Approach type	Algorithm	Accuracy
<b>01</b>	Deep Learning	ANN	91%
<b>02</b>	Tree based	Decision Tree	60%
<b>03</b>	Ensemble	Random Forest	51%

Table 1.3: Comparison with other models

The error rate of our model is very low. 91% times it was giving the accurate answer. We have tested our model with some similar kind of data by which the model was trained.

With a little bit of modification, we prepared the test data of 1431 texts. For 1302 times it generated the notes correctly. And 129 times it did not show the desired output. Figure 2.2 explains this result.

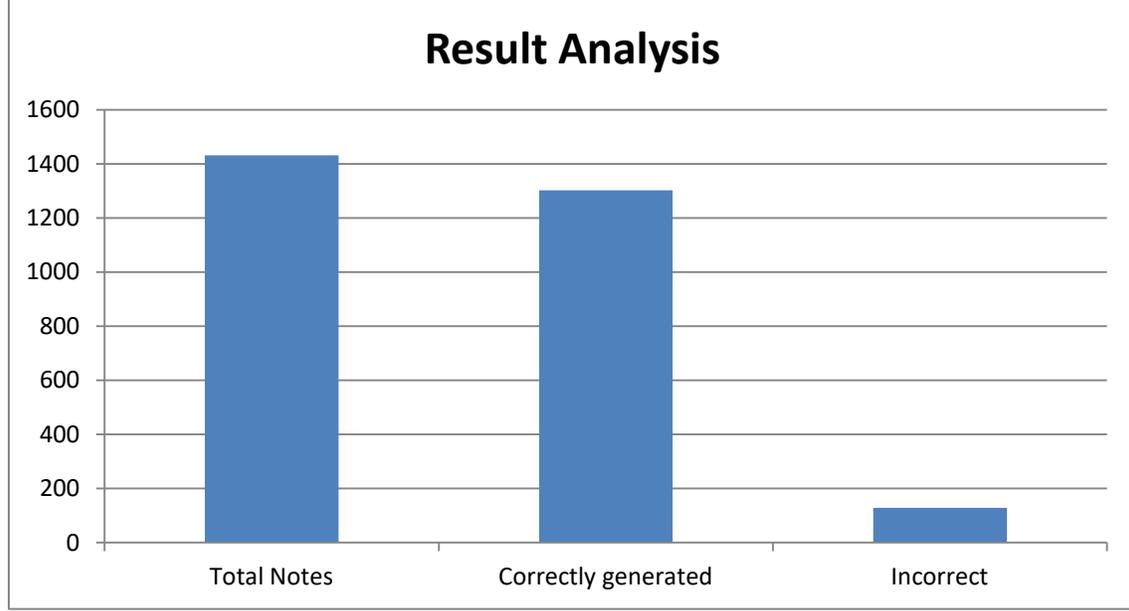


Figure 2.2: Analysis of results

After finishing the experiment we created a pipeline by which we were able to give input to the model and it was generating large notes based on the pattern match. In table 1.4-1.8 the demo of the given input and output is shown briefly.

Actual pattern	ইসলামবাগে প্লাস্টিক কারখানায় আগুন
Note	রাজধানীর পুরান ঢাকার ইসলামবাগে একটি প্লাস্টিক পণ্য তৈরির কারখানায় আগুন লেগেছে। আজ বৃহস্পতিবার রাত ৯টা ৫৫ মিনিটে বাগানবাড়ি এলাকায় এই অগ্নিকাণ্ডের ঘটনা ঘটে। ফায়ার সার্ভিসের নিয়ন্ত্রণকক্ষে দায়িত্বরত কর্মকর্তা ফরহা দুজ্জামান প্রথম আলোকে জানান, আগুন নেভাতে ফায়ার সার্ভিসের পাঁচটি ইউনিট কাজ করছে। তবে আগুন লাগার কারণ এখনো জানা যায়নি।
Given input	ইসলামবাগে প্লাস্টিক
Generated Note	রাজধানীর পুরান ঢাকার ইসলামবাগে একটি প্লাস্টিক পণ্য তৈরির কারখানায় আগুন লেগেছে। আজ বৃহস্পতিবার রাত ৯টা ৫৫ মিনিটে বাগানবাড়ি এলাকায় এই

	অগ্নিকাণ্ডের ঘটনা ঘটে।ফায়ার সার্ভিসের নিয়ন্ত্রণকক্ষে দায়িত্বরত কর্মকর্তা ফরহা দুজ্জামান প্রথম আলোকে জানান, আগুন নেভাতে ফায়ার সার্ভিসের পাঁচটি ইউনিট কাজ করছে। তবে আগুন লাগার কারণ এখনো জানা যায়নি।
--	---

Table 1.4: Input and output sample 1

Actual pattern	করোনাভাইরাসে প্রচুর মারা গিয়েছে
Note	প্রসঙ্গত প্রাণঘাতী করোনাভাইরাসে আক্রান্ত হয়ে যুক্তরাষ্ট্রে গত ২৪ ঘণ্টায় এক হাজার ১৬৯ জনের মৃত্যু হয়েছে। বৃহস্পতিবার জনস হপকিন্স বিশ্ববিদ্যালয়ের ট্রাকারে এমন তথ্যই দেখিয়েছে।বৈশ্বিক মহামারীটি শুরু হওয়ার পর কোনো দেশে একদিনে এটিই সর্বোচ্চ মৃত্যুর সংখ্যা।বুধবার রাত সাড়ে ৮টা থেকে পরদিন একই সময় পর্যন্ত এসব মানুষের মৃত্যু হয়েছে।এর আগে গত ২৭ মার্চ ইতালিতে একদিনে ৯৬৯ জনের মৃত্যু হয়েছে। করোনাভাইরাসে যুক্তরাষ্ট্রে এখন পর্যন্ত পাঁচ হাজার ৯২৬ জনের মৃত্যু হয়েছে আক্রান্তের সংখ্যায় দেশটি আগে থেকেই সবার চেয়ে এগিয়ে ছিল। বৃহস্পতিবার শেষ খবর পাওয়া পর্যন্ত এ সংখ্যা দুই লাখ ৪৫ হাজার ছাড়িয়ে।
Given input	করোনাভাইরাসে গিয়েছে
Generated Note	প্রসঙ্গত প্রাণঘাতী করোনাভাইরাসে আক্রান্ত হয়ে যুক্তরাষ্ট্রে গত ২৪ ঘণ্টায় এক হাজার ১৬৯ জনের মৃত্যু হয়েছে। বৃহস্পতিবার জনস হপকিন্স বিশ্ববিদ্যালয়ের ট্রাকারে এমন তথ্যই দেখিয়েছে।বৈশ্বিক মহামারীটি শুরু হওয়ার পর কোনো দেশে একদিনে এটিই সর্বোচ্চ মৃত্যুর সংখ্যা।বুধবার রাত সাড়ে ৮টা থেকে পরদিন একই সময় পর্যন্ত এসব মানুষের মৃত্যু হয়েছে।এর আগে গত ২৭ মার্চ ইতালিতে একদিনে ৯৬৯ জনের মৃত্যু হয়েছে। করোনাভাইরাসে যুক্তরাষ্ট্রে এখন পর্যন্ত পাঁচ হাজার ৯২৬ জনের মৃত্যু হয়েছে আক্রান্তের সংখ্যায় দেশটি আগে থেকেই সবার চেয়ে এগিয়ে ছিল। বৃহস্পতিবার শেষ খবর পাওয়া পর্যন্ত এ সংখ্যা দুই লাখ ৪৫ হাজার ছাড়িয়ে।

Table 1.5: Input and output sample 2

Actual pattern	শ্যামগঞ্জের করোনা ঠেকাতে মাঠে লাল পোশাকধারী স্বেচ্ছাসেবক
----------------	--

Note	ময়মনসিংহের শ্যামগঞ্জ করোনাভাইরাস মোকাবেলায় দিন-রাত শ্রম দিয়ে যাচ্ছেন লাল রঙের পোশাক পরা স্ব্চ্ছাসেবীরা। গত ৬ দিন ধরে নিজ উদ্যোগে এলাকায় জীবাণুনাশক স্প্রে করা সহ করোনাভাইরাস নির্মূলে নানা উদ্যোগ হাতে নিয়েছেন এসব স্ব্চ্ছাসেবী। লাল রঙের পোশাকের কারণে স্থানীয়রা তাদের লালবাহিনী নাম দিয়েছে। সামাজিক মাধ্যম ফেসবুকেও এ লালবাহিনীর সামাজিক কর্মের ছবি ভাইরাল হয়েছে, যা বেশ প্রশংসিত হচ্ছে। জানা গেছে, নিজের তাগিদেই প্রথমে কাজ শুরু করেন শ্যামগঞ্জ উচ্চ বিদ্যালয়ের সহকারী শিক্ষক গোবিন্দ বণিক। তাকে সহযোগিতায় এগিয়ে আসেন স্থানীয় সংগঠন স্বপ্ন সমবায় সমিতি।
Given input	শ্যামগঞ্জ মাঠে লাল পোশাকধারী স্ব্চ্ছাসেবক
Generated Note	ময়মনসিংহের শ্যামগঞ্জ করোনাভাইরাস মোকাবেলায় দিন-রাত শ্রম দিয়ে যাচ্ছেন লাল রঙের পোশাক পরা স্ব্চ্ছাসেবীরা। গত ৬ দিন ধরে নিজ উদ্যোগে এলাকায় জীবাণুনাশক স্প্রে করা সহ করোনাভাইরাস নির্মূলে নানা উদ্যোগ হাতে নিয়েছেন এসব স্ব্চ্ছাসেবী। লাল রঙের পোশাকের কারণে স্থানীয়রা তাদের লালবাহিনী নাম দিয়েছে। সামাজিক মাধ্যম ফেসবুকেও এ লালবাহিনীর সামাজিক কর্মের ছবি ভাইরাল হয়েছে, যা বেশ প্রশংসিত হচ্ছে। জানা গেছে, নিজের তাগিদেই প্রথমে কাজ শুরু করেন শ্যামগঞ্জ উচ্চ বিদ্যালয়ের সহকারী শিক্ষক গোবিন্দ বণিক। তাকে সহযোগিতায় এগিয়ে আসেন স্থানীয় সংগঠন স্বপ্ন সমবায় সমিতি।

Table 1.6: Input and output sample 3

Actual pattern	তাবলিগের ৯৬০ বিদেশি মুসল্লিকে কালো তালিকাভুক্ত করল ভারত
Note	করোনাভাইরাসের সংক্রমণের মধ্যেই দিল্লির তাবলিগ জামাত সম্মেলনে যোগ দেয়া ৯৬০ জন বিদেশি মুসল্লিকে কালো তালিকাভুক্ত করেছে ভারত। দেশটির কেন্দ্রীয় স্বরাষ্ট্র মন্ত্রণালয় জানিয়েছে, ৬৭ দেশের ওই ৯৬০ জনের ভারতীয় ভিসা বাতিল করা হয়েছে। তাদের দেশে ফেরত পাঠানোর পাশাপাশি আইনি ব্যবস্থা নেয়া হবে। ভবিষ্যতে তাবলিগ জামাতে অংশগ্রহণ করতে চেয়ে ভিসার আবেদন করলে তাদের আর ভিসা দেয়া হবে না বলেও জানানো হয়েছে। মন্ত্রণালয় সূত্রের বরাতে আনন্দবাজার পত্রিকা জানিয়েছে, মার্কিন যুক্তরাষ্ট্র, ইতালি, ফ্রান্স সহ বিভিন্ন দেশের প্রায় ১৩০০ প্রতিনিধি যোগ দিয়েছিলেন তাবলিগ জামাতে। সেসব প্রতিনিধি এবং তাদের সংস্পর্শে আসা মিলিয়ে বৃহস্পতিবার পর্যন্ত মোট ৫৫০ জনের কোভিড-১৯ পজিটিভ রিপোর্ট এসেছে।

Given input	তাবলিগের মুসল্লিকে কালো তালিকাভুক্ত করল ভারত
Generated output	করোনাভাইরাসের সংক্রমণের মধ্যেই দিল্লির তাবলিগ জামাত সম্মেলনে যোগ দেয়া ৯৬০ জন বিদেশি মুসল্লিকে কালো তালিকাভুক্ত করেছে ভারত। দেশটির কেন্দ্রীয় স্বরাষ্ট্র মন্ত্রণালয় জানিয়েছে, ৬৭ দেশের ওই ৯৬০ জনের ভারতীয় ভিসা বাতিল করা হয়েছে। তাদের দেশে ফেরতপাঠানোর পাশাপাশি আইনিব্যবস্থা নেয়া হবে। ভবিষ্যতে তাবলিগ জামাতে অংশগ্রহণ করতে চেয়ে ভিসার আবেদন করলে তাদের আর ভিসা দেয়া হবে না বলেও জানানো হয়েছে। মন্ত্রণালয় সূত্রের বরাতে আনন্দবাজার পত্রিকা জানিয়েছে, মার্কিন যুক্তরাষ্ট্র, ইতালি, ফ্রান্সসহ বিভিন্ন দেশের প্রায় ১৩০০ প্রতিনিধি যোগ দিয়েছিলেন তাবলিগ জামাতে। সেসব প্রতিনিধি এবং তাদের সংস্পর্শে আসা মিলিয়ে বৃহস্পতিবার পর্যন্ত মোট ৫৫০ জনের কোভিড-১৯ পজিটিভ রিপোর্ট এসেছে।

Table 1.7: Input and output sample 4

Actual pattern	চৌগাছায় চাচাতো ভাইদের হামলায় আহত পরিবহনচালক নিহত
Note	যশোরের চৌগাছায় জমি নিয়ে দ্বন্দ্ব চাচাতো ভাইদের হামলায় আহত মুক্তার হোসেন (৫৫) নামে এক পরিবহনচালক মারা গেছেন। শুক্রবার ভোরে ঢাকা মেডিকেল কলেজ হাসপাতালে চিকিৎসাধীন তার মৃত্যু হয় নিহত মুক্তার হোসেন উপজেলার মসিয়ুরনগর গ্রামের জাহাবুল্লের ছেলে। তিনি চৌগাছা-ঢাকা রুটে চলাচলকারী রয়েল পরিবহনের চালক ছিলেন।
Given input	হামলায় আহত পরিবহনচালক নিহত
Generated Note	যশোরের চৌগাছায় জমি নিয়ে দ্বন্দ্ব চাচাতো ভাইদের হামলায় আহত মুক্তার হোসেন (৫৫) নামে এক পরিবহনচালক মারা গেছেন। শুক্রবার ভোরে ঢাকা মেডিকেল কলেজ হাসপাতালে চিকিৎসাধীন তার মৃত্যু হয় নিহত মুক্তার হোসেন উপজেলার মসিয়ুরনগর গ্রামের জাহাবুল্লের ছেলে। তিনি চৌগাছা-ঢাকা রুটে চলাচলকারী রয়েল পরিবহনের চালক ছিলেন।

Table 1.8: Input and output sample 5

## CHAPTER 5

### Conclusion and future work

Our desired model generated the notes quite efficiently. Hence no model is 100% accurate when you have to deal with large dimensional text data. But in the case of our issue, it was showing quite good results and achieved an outstanding accuracy of 91% which is huge when the dataset is related with complex text patterns. From data collection to determine the final output of this research study it took around 6-7 months of hardworking. Still, there were some limitations which we could not solve.

The experiment would have been more accurate if we could have provided more data. But due to lack of time and funding we could not include huge dataset for our experiment. There was also a lack of experimental tools for Bengali text analysis. Though Bengali NLP researchers are trying their best to introduce new resources, still there is a lot of diversity we need to enrich the Bengali language. Good libraries like NLTK or Spacy do not provide support of Stemming, Lemmitization, stop-words, etc for the Bengali language. Hence we had to do it by raw coding. And appropriate pre-trained word embedding was also not available for the Bengali language. That's why we couldn't perform word similarity measurement, and for this issue, our model was also skipping this part. LSTM could have been used if we could have solved this issue. There are some works which we could not complete due to lack of time but willing to do it in the future.

We will try to implement this study in a real-life application. Tools like automated software in the Bengali language will add an important resource in "Bangla NLP". If possible we will train this model with more and more data so that the model performs better than before.

#### Reference:

- [1] Palm G. (1986) Warren McCulloch and Walter Pitts: A Logical Calculus of the Ideas Immanent in Nervous Activity. In: Palm G., Aertsen A. (eds) Brain Theory. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-70911-1\\_14](https://doi.org/10.1007/978-3-642-70911-1_14)

- [2] <https://www.sciencedirect.com/topics/engineering/backpropagation-learning>
- [3] Hochreiter, S. and Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation*, 9(8), pp.1735-1780. [4] Weizenbaum, J.: Contextual understanding by computers. *Commun. ACM* 10(8), 474–480 (1967)
- [5] Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. *Commun.ACM* 9,36–45(1966)
- [6] ALICE:Artificial Intelligence Foundation. [http://www.alicebot.org/bios/rechard wallace.html](http://www.alicebot.org/bios/rechard-wallace.html)
- [7] Jabberwacky. <http://www.jabberwacky.com>
- [8] Bickmore, T., Schulman, D., Vardoulakis, L.: Tinker - a relational agent museum guide. *J. Auton. Agents Multi Agent Syst.* 27(2), 254–276 (2013).
- [9] Bickmore, T., Schulman, D.: Practical approaches to comforting users with relational agents. In: *ACM SIGCHI Conference on Human Factors in Computing Systems* (2007).
- [10] Cassell, J.: Embodied conversational agent: representation and intelligence in user interfaces. *AI Mag.* 22(4), 67–83 (2001)
- [11] G. K. Vamsi, A. Rasool and G. Hajela, "Chatbot: A Deep Neural Network Based Human to Machine Conversation Model," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225395.
- [12] M. Kowsher, F. S. Tithi, M. Ashraful Alam, M. N. Huda, M. Md Moheuddin and M. G. Rosul, "Doly: Bengali Chatbot for Bengali Education," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934592.
- [13] M. Kowsher, M. A. Alam, M. J. Uddin, M. R. Islam, N. Pias and A. R. M. Saifullah, "Bengali Informative Chatbot," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2019, pp. 1-7, doi: 10.1109/IC4ME247184.2019.9036585.
- [14] M. Keya, A. K. M. Masum, B. Majumdar, S. A. Hossain and S. Abujar, "Bengali Question Answering System Using Seq2Seq Learning Based on General Knowledge Dataset," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225605.
- [15] [https://mlcheatsheet.readthedocs.io/en/latest/activation\\_functions.html](https://mlcheatsheet.readthedocs.io/en/latest/activation_functions.html)
- [16] <https://seongjuhong.com/2019-12-05pm-softmax-function/>
- [17] YoungGang Cao [a,1] , Feifan Lia [a], Pippa Simpson [b], Lamont Antieau [a], Andrew Bennett [c,d J. James Cimino [e], John Ely [f], Hong Yu [a,g,\*]: AskHERMES: An online question answering system for complex clinical questions.( *Journal of Biomedical Informatics* 44 (2011) 277–288)
- [18] Somnath Banerjee, Sivaji Bandyopadhyay: Bengali Question Classification: Towards Developing QA System.( *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP)*, pages 25–40, COLING 2012, Mumbai, December 2012.)
- [19] L. HIRSCHMAN (a1) and R. GAIZAUSKAS (a2): Natural language question answering: the view from here.( *Published online by Cambridge University Press: 14 February 2002*)

[20] Chandra Obula Reddy [1], Dr. K. Madhavi [2]: A survey on Types of Question Answering system. (IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 6, Ver. IV (Nov.- Dec. 2017), PP 19-23)

## A DEEP LEARNING APPROACH FOR GOVERNMENT AUTOMATED BENGALI NOTE WRITING SERVICE

### ORIGINALITY REPORT

<b>3%</b> SIMILARITY INDEX	<b>2%</b> INTERNET SOURCES	<b>2%</b> PUBLICATIONS	<b>1%</b> STUDENT PAPERS
-------------------------------	-------------------------------	---------------------------	-----------------------------

### PRIMARY SOURCES

<b>1</b>	<b>www.coursehero.com</b> Internet Source	<b>2%</b>
<b>2</b>	<b>link.springer.com</b> Internet Source	<b>&lt;1%</b>
<b>3</b>	<b>Submitted to University of Westminster</b> Student Paper	<b>&lt;1%</b>
<b>4</b>	<b>Submitted to Kennedy-Western University</b> Student Paper	<b>&lt;1%</b>
<b>5</b>	<b>Hasina Attaullah, Faiza Iqbal, Muhammad Younus Javed. "Intelligent vertical handover decision model to improve QoS", 2008 Third International Conference on Digital Information Management, 2008</b> Publication	<b>&lt;1%</b>
<b>6</b>	<b>www.mdpi.com</b> Internet Source	<b>&lt;1%</b>
<b>7</b>	<b>manip.crhc.uiuc.edu</b> Internet Source	<b>&lt;1%</b>