SHAHJALAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

AUTOMATED QUESTION-ANSWERING SYSTEM FOR ADMISSION TEST HELPINE IN BANGLADESH

Aundrila Acharjee[1], Dayamoy Datta Saikat[2]
2015331039[1], 2015331058[2]

January 15, 2019

Abstract

We have made an automate Question-Answering System.Here, it tries to give answer that are asked by the examinee who are going to take participate in admission test.At first we have collected a good amount of data and have analyzed them by different charts and graphs. Then we have used different methodologies to train the chatbot that is explained.Then we have shown our system architecture how does it work.At last,we have shown our accuracy.

Key words: Dataset, Data Analysis, Train Chatbot, Accuracy

1 Introduction

In Bangladesh, all the public and engineering universities held a Admission test each year .Till now, Admission helpdesks have been running manually.Students ask questions manually and all the helpdesks answer that questions manually also. We are trying to implement a **Question-Answering System** using **Natural Language Processing(NLP)** that will answer all the asked questions automatically.

Natural language processing (NLP) is a branch of **Artificial Intelligence** that helps computers understand, interpret and manipulate human language.So far, less works have been done on Bangla Language.Hence we are trying to implement the whole **Question-Answering System** based on Bangla Language.

In admission test,more or less all the candidates have so many questions on various topics like - preparation process, different queries on universities, exam time and schedule, subject choice and so on. Generally they ask all these questions in different admission helpdesks.People knowing the answer, answers the questions. Sometimes the answers might be right or wrong. Or sometimes they don't get any answer.
We are going to build a automated system for this purpose.Our system will cover admission before-after queries.The system will answer all the question using **Machine Learning**.
The main challenge on this project is a good data-set. Previously,no remarkable works have been done on data-set. Hence our main concern is to build a good data-set.

The upcoming sectors of this report covers the properties and classification of the data-set, positive and negative sides of the automated system, performance of our running system, previous works regarding this system ,result of the system,conclusion and discussion.

2 Data Analysis

We have collected around 1000 data.
We collected data manually from various facebook groups such as SUST Admission Helpline,BUET Admission helpline, DU admission helpdesk etc. and websites like sust.edu, admissionwar, buet.ac.bd etc.
We have divided our dataset into three parts:
1.Bangla
2.Banglish
3.English
We have around 50% Bangla and 45% Banglish and 5% English data.
We have collected following universities data:

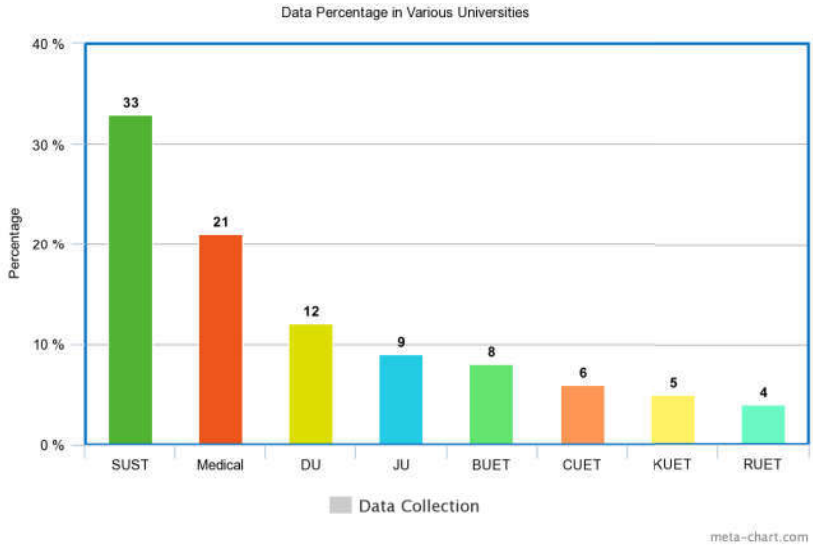

Figure 2.1: University data collection

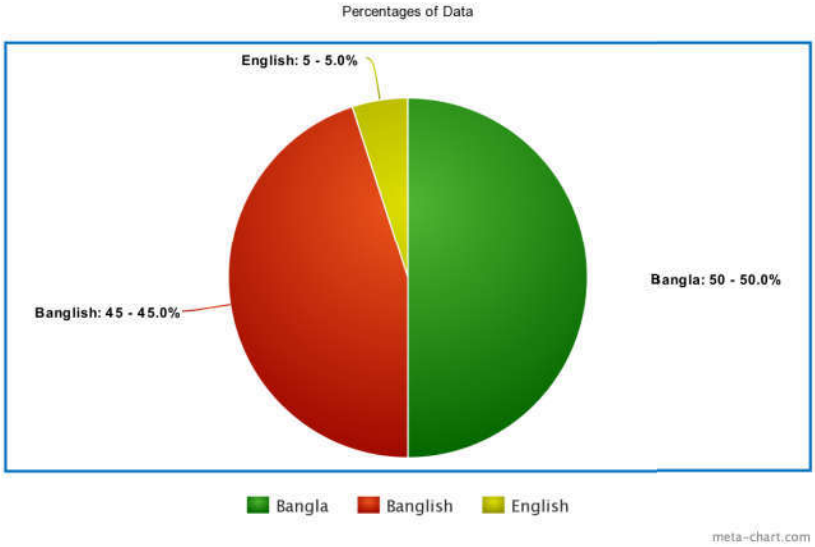

Figure 2.2: Data collection percantage based on language

We formatted our data in CSV(Comma Separated Value) file.

In our CSV file three types of tag we used:

1.Question
2.Answer
3.University
4.Language

3 Methodology

We tried to use two types of methods to train our data-set
i) Chatfuel API[1]
ii) Chatterbot Module[2]
In Chatfuel API, data is about to be trained one by one. So,it's very tough to train our data-set onto this API. Though it is tough, we trained few of our data manually and it worked almost fine. Then we switched this method and started to use chatterbot corpus module to train our data. Within 10 asked question it almost can answer 4-5 questions correctly. So we can say that this model accuracy is about 40-50% . It wasn't so good because used classifier in chatterbot is very naive.

4 System Design

Earlier in the Methodology section, we claim that we used chatterbot module to train our data. On this module, their is a classifier called list trainer. We trained our data-set by list trainer classifier. It is very naive classifier.

4.1 How ChatterBot Works

An untrained instance of ChatterBot starts off with no knowledge of how to communicate. Each time a user enters a statement, the library saves the text that they entered and the text that the statement was in response to. As ChatterBot receives more input the number of responses that it can reply and the accuracy of each response in relation to the input statement increase.

The program selects the closest matching response by searching for the closest matching known statement that matches the input, it then chooses a response from the selection of known responses to that statement.

4.2 Process flow diagram

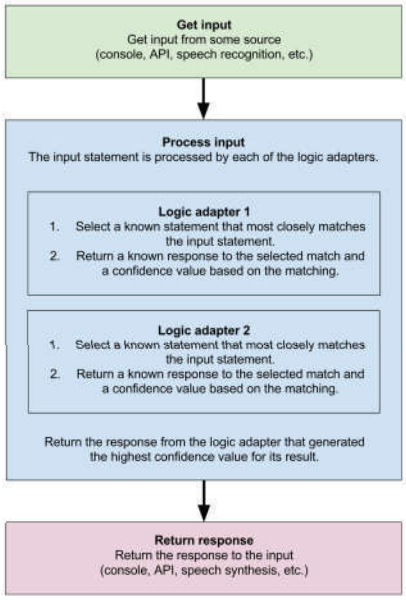

Figure 4.1: Process flow diagram

5 Development

We didn't implement our chatbot in any platform. We just run a python code importing chatterbot model for checking the result. By using more preprocessing of data-set,enlarging dataset with variety and NLTK,we can make this chatbot better. Actually it will be helpful for student,tutor and guardian who are searching many queries about admission helpline.

6 Result

We trained our data-set into naive bayes classifier and SVM classifier with a simple python code to measure our performance.

Then, we split 70% data into training set and 30% into test set.

In SVM,Accuracy was 44.379%
In Naive Bayes, Accuracy was 36.345%

Between this two SVM was shown better result for our data-set. This performance can be increased by data preprocessing and enlarging data-set.

7 conclusion

In short, we have made an automated Q/A System with 40-50% accuracy. Its' main concern is to help the examinee by providing accurate answers to their asked questions.

References

[1] Chatfuel. https://dashboard.chatfuel.com/. Accessed: 2019-01-15.

[2] Chatterbot. https://chatterbot.readthedocs.io/en/stable/. Accessed: 2019-01-15.