Doly: Bengali Chatbot
for Bengali Education

1st Md. Kowsher
Applied Mathematics
Noakhali Science and Technology
University, Noakhlai-3814
Noakhlai,Bangladesh.
ga.kowsher@gmail.com

2nd Farhana Sharmin Tithi
Computer science and engennering
Daffodil International University,
Dhaka,Bangladesh
farhanasharmin3tithi@gmail.com

3rd M Ashraful Alam
Applied Mathematics
Noakhali Science and Technology
University, Noakhlai-3814
Noakhlai,Bangladesh.
mdashrafulsikdar1580@gmail.com

4th Mohammad Nurul Huda
MSCSE
United International University,
United city,Madani Avenue,Badda.
Dhaka 121,Bangladesh.
mnh@cse.uiu.ac.bd

5st Mir Md Moheuddin
Applied Mathematics
Noakhali Science and Technology
University, Noakhlai-3814
Noakhlai,Bangladesh.
mirmdmoheuddinkhan@gmail.com

6th Md. Golam Rosul
Applied Mathematics
Noakhali Science and Technology
University, Noakhlai-3814
Noakhlai,Bangladesh.
golamrosul1506040@gmail.co

Abstract—: **This Scientific Research paper is a procedure of an automated system "Doly: Bengali Chatbot" which gives a reply to a user query on behalf of a human for the education system in the Bengali language. This is an AI-based Chatbot, mainly based on machine learning algorithms and Bengali Natural Language Processing (BNLP). The machine gets embedded with this knowledge to identify the desired sentences and making a decision within itself, as a response to answer questions. There are many English Chatbot's which used in education, web query, banking sector & various sectors. In this research, we have propounded a complete data-driven retrieval based closed domain Chatbot which is easily colloquy in the Bengali language with the users. We've created the train function adapter to train the Doly by encoding (encoding="utf8") our corpus from bot data. An input adapter has been created to take input and for output, an output adapter has been created to generate automated responses to a user's input. We have also used a machine learning algorithm like search algorithm for finding an appropriate list of matching results from the corpus and use Naïve Bayesian algorithm to generate the right answer from data. The main aim of this Chatbot based system is to bridge the gap between the knowledge sources by providing instant replies to the questions and queries that have to ask in the Bengali language.**

Index terms*:* Chatbot, Dialogue & Corpus, Naïve Bayesian algorithm, AI, Bengali Natural Language Processing (BNLP).

I. INTRODUCTION

By Technology-based education, learners can easily get access to unlimited online resources. It creates huge facilities for technical learning, distance learning, and self-learning. Nowadays chatbot is using as an only one learning platform for the education system. Personal education chatbot will help students to interest by self-learning & also it can also remove the dependency from tutor & other technology. The chatbot is a conversational intelligent agent that can easily communicate with human users by using Natural Language Processing. It is likewise known as a chatterbot, Bot, IM bot, talkbot, Interactive agent, or Artificial Conversational Entity. The user can interact with the bot in written, oral, or mixed format of conversation.

The applications of chatbots can be applied in various fields such as Health, Tourism Business,Customer Confirmation and so on. The architecture of Chabot integrates language models with computational algorithms. Normally, a chatbot works by a user asking a question or developing a comment and replying the question or initiating a new topic. We can deliver instance examples of Cleverbot [1] ALICE [1] etc.The Bengali language is the main language of Bangladesh and the fourth most popular language in the world. Bangla is spoken by about 245 million people of Bangladesh and two states of India [3]. The literacy rate of Bangladesh is 75.4% and the numbers of total students are 23,907,151. The rate of literacy is increasing day by day in Bangladesh. Nowadays the educations system is being dependent on Intelligent Instructing Systems of computer circumstance for education.This is why the paper proposes the Bengali language chatbot named 'Doly' based on the Bengali language with the help of Natural Language Processing (NLP) as well as Machine Learning (ML) techniques for developing an education tool of Bangladesh. The trial spectacle is that our chatbot is capable to impart responses to the users in the real time. To constructing Bengali chatbot 'Doly', we installed Python 3.7.0 & Anaconda in our Machine. For Natural Language Processing (NLP) we installed Natural Language Toolkit (NLTK). We also need to install Chatterbot that comes with built-in adapter classes that allow it to connect to various types of databases. We can't run directly the Bengali chatbot in chatterbot as it did not support Unicode.As a Unicode Decode Error for the implementation of Bengali chatbot 'Doly', we have to encode Bangla NLP with a chatterbot. We created the train function adapter to train the Doly by encoding our corpus from bot data. It this case we use encoding="utf8" in our .yml corpus to read Bengali data. To take input we created an input adapter and for output, we created an output adapter to generate automated responses to a user's input. This makes it easy to create chatbots and automate conversations with users. The task separated by four parts, training data, input data, process input, and return response or output. We also use machine learning algorithm

like search algorithm for finding the appropriate list of matching results from the corpus and also use a Bayesian algorithm to select right answer form list of matching result.This paper described a created chatbot that returns a relevant answer to questions asked by the user in Bangla language.

The contributions are summarized as follows:

- We have proposed a Bengali chatbot as 'Doly' by imparting Bengali Natural Language Processing.
- For the pre-processing Bengali language, we have created our own BLTK tool in python.
- As algorithms, we have used Naive Bayes classifier and popular algorithm Edit distance of dynamic programming
- Various adapters were developed for our proposal task.

The paper is composed as follows, Section 2 gives the details of the related works on the chatbot technology. Section 3 contains the Bengali Natural Language Processing (BNLP). Section 4 described the design and methods used for Creating Doly: Bengali Chatbot. In section 5 offers the results and analysis. Section 6 included Comparison between others Chatbot. Section 7 concludes research work with future directions. Finally, Section 8 highlights the conclusion.

II. RELATED WORK

Weizenbaum, J. et al.[1] discussed the first idea of chatbots originated in the Massachusetts Institute of Technology [2] where the Eliza chatbot (also known as "Doctor"). The users of this chatbot imagined Computer programs really could hear their problems and understand, and help them in a constructive way. There are other chatbots called ALICE that used heuristic pattern matching to follow the rules from the input [3].Breazeal, C. et al.[4] research has a robotic chat-bot which understand and relate to users in human or social terms, adapt and learn throughout their lifetime. It also communicates with users. Vinyals, O. et al.[5] discuss the converse model in natural language interaction that predicting the very next command given the previous command in conversations. Bickmore et al.[6] researched agent based chatbot. This chatbot is relational computer agents that they mapping social-emotional relationships for users. An artificial intelligence chatbot like Jabberwacky [7], that studies language and context-based human interactions. To find appropriate response to some stored data such as conversations and comments.Bickmore et al.[8] also research a relational agent which applied different relational behaviors to establish social bonds with museum visitors and evaluate the ability of an experimental relational agent to comfort users in stressful situations [9]. Cassell, J. et al.[10] The most popular agent called Rea that shows non-verbal behaviors such as hand gestures, facial displays, an eye gaze, body posture etc. This behaviors improve the effectiveness of communication and can create a more immersive experience for users. Augello et al.[11] proposed a social chatbot model for a communicative skill learning game based on "Social practice" that can choose the most appropriate dialogue. Hill J. et al.[12] shows a different chatbot such as human-chatbot conversations lack in content, quality and

vocabulary and human-human conversations over IM.Pan, X. et al.[13] discussed a virtual agents chatbot. They suggested a projection based system virtual party like an environment. In this system, a conversational male user approached by a female character called Christine. Christine also can formulate personal questions and statements. This female character is interested in a male user that involves smiles, headnotes, eye contact, leaning towards. Crutzen, R. et al.[14]studies an information chatbot that helps users to know about sex, drugs, and alcohol use among adolescents. This chatbot also called question related chatbot or search engines. Shawar et al.[15] described existing work on chatbot knowledge acquisition is mainly based on human annotated datasets. Their approaches are helpful to develop sense knowledge for Bangla chatbots, but there are not available of extracting knowledge for the specific task and technical challenges and design issues in Bangla language processing[16]. Shrestha et al. [17] proposed a way is to detect pairs in an email conversation for the work of email summarization. Zhou and Hovy [18] developed a summarization system for technical chats and emails about the Linux kernel.

Unlike these works, we have created a search based chatbot for Bengali language using BNLP. For Bengali Natural Language processing, we have created BLTK and BRE tools.

III. BENGALI NATURAL LANGUAGE PROCESSING (BNLP)

The Bengali language is the fourth most spoken language in the world. It is the mother tongue of the people of Bangladesh; there are also few parts of India like as Calcutta who talk with Bangla. But it is really a matter of great mournful that the Bengali language deliberated as the low-density language because of the insufficient digitized text element is available in the Bengali language. But nowadays Bengali Natural Language Processing (BNLP) is the most interesting research field and AI (NLP) filed in Bangladesh. Developing an ideal Chatbot is mandatory because it needs a huge database of Bangla Language and must give reasonable answers to all interactions. Due to lack of availability of large Bengali conversation corpus, for the prototype, the system enhanced with a sample hand-crafted dataset for basic interactions. So the main goal of our task is to develop an ideal chatbot in Bengali which will able to interact in Bangla Language as well as building a corpus in Bangla. We name it "Doly".

বাংলা স্বরবর্ণ
অ আ ই ঈ উ ঊ ঋ এ ঐ ও ঔ
বাংলা ব্যঞ্জনবর্ণ
ক খ গ ঘ ঙ   প ফ ব ভ ম
চ ছ জ ঝ ঞ   য র ল শ ষ
ট ঠ ড ঢ ণ   স হ ড় ঢ় য়
ত থ দ ধ ন   ৎ ং ঃ ঁ

Fig. 2. Bengali Alphabat.

IV. PROPOSED WORK

We have introduced a simple encoder and decoder based conversational method of an intelligent agent that will provide for chatbot users an existence with Bengali knowledge base by interactively asking for its characteristic.

In the section of BNLP helps to encode or decode Bengali corpus for training, input, and output return. BNLP is also imparted for the pre-processing questions and answer.

In the beginning, 'Doly' is nothing but an unskillful chatbot which introduction herself with no sense of how to connect with conversations. In a certain moment a user inputs affirmation, the store protects the affirmation that they are given and the affirmation that was in reply. As 'Doly' accepts more input the number of affirmation protected in the library and that it can reply and the accuracy of each answer increase the relation with the input affirmation.

The system chooses the best closest matching reply by investigation the closest matching acquainted statement that matches the input; it then chooses an answer from the chosen of acquainted reply to that affirmation.
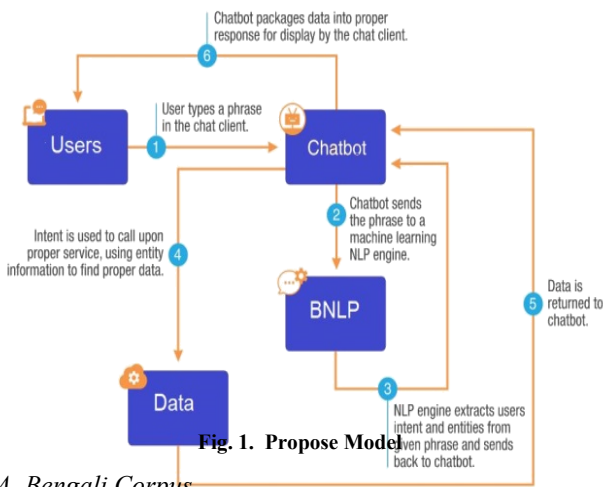
Fig. 1. Propose Model

A. Bengali Corpus

For dialog corpus of Bengali chatbot Doly we have made some simple Bengali Language corpus these are saved by .yml or yaml file. YAML is an acronym that keeps for "YAML Ain't Markup Language" and is very immensely common, intelligent, human-friendly and ideal for data serialization for all programming languages. To complete our novel task, we collected 2393 questions and its related answer as our inserted data of Bengali educations. We separated the whole data as their related group.

Fig. 3. Bengali Corpus use in Doly Chatbot.

B. Environmental setup

Natural Language Processing method and Natural Language Toolkit for Python are provided developing an apparatus in order to explore speech, intelligent explanation and human-like responses. While constructing of Bengali chatbot 'Doly', Python 3.7.0 is used in our Machine. Python is a high-level object-oriented language (OOP) and is suitable for scientific examination, tools development. We have also used the Anaconda as the apportionment of Python. Anaconda creates the best stage for open source data science which is powered by Python. For Natural Language Processing (NLP) we installed Natural Language Toolkit (NLTK). We also need to install ChatterBot that comes with built-in adapter classes that allow it to connect to various types of databases.

Python 3.7.0 is used for the implementation of Bengali chatbot 'Doly' because of providing "Unicode Decode Error". Unicode Decode Error is a runtime error that is caused by non-English language like Bengali with a large number of letters in the alphabet. The Unicode range of Bengali is 0980–09FF, it has 11 vowels and 40 consonants. To encode the Bengali Language, the encoding="utf8" is used that has consonant conjuncts, modifier, and other graphemes. So Bengali cannot be set up with ASCII decoding system. In many pre-processing steps, NLTK has no value for the Bengali Language such as to remove a stop word, lemmatization etc. In this case, we have created BLTK for the preprocessing all inserted questions and answer.

Fig. 4. Official Unicode Consortium Code Chart of Bengali.

C. Procedure

We have created the train function adapter to train the Doly by encoding our corpus from bot data. The bot data are encoded using encoding ="utf8" in our .yml corpus to read Bengali data. For input and output, the input and for output adapter are created. The output adapter results the response of user question and deals with response adapter that is generating the automated responses. This makes it easy to create chatbots and automate conversations with users. The task divided into four parts, training data, input data, process input and return response or output.

1) Pre-Processing

To prepare data for the training, we need to pre-process the whole dataset. Various types of pre-processing methods are applied such as anaphora, removing stop words and lemmatization. 'Anaphoric' which describe to a word that is led in a sentence and replacing pronoun from a noun. It needs a successful identification and resolution of NLP. In the proposed chatbot, we have limned a review of the task

done in the field of anaphora resolution in the Bengali language that is the influence of cases on pronouns, mainly personal pronoun. The Hobbs' algorithm of Anaphora resolution have been used in our chatbot system. Another preprocessing technique is to remove stop words. A stop word removing is an action which commonly applied to delete a word that does not impress on documents or sentences for an instance of Bengali stop words. The other preprocessing process is lemmatization. Lemmatization is an efficient process to turn a word to its root word. It's one kind of NLP technique and applied for imparting in several purposes such as text mining, questions replying or Chatbot and others that boost to deal with root words of any language. In our proposed Chatbot, Two novel techniques such as DBSRA and Trie are used for lemmatization based on data structure and the hash table of computer programming. 'DBSRA' refers to 'Dictionary Based Search by Removing Affix' which is a really easy concept and more suitable for the lemmatization of Bengali words with the lowest time and space complexity.

2) Training Via Corpus Data

This training action authorizes to 'Doly' to be trained through a catalog of strings where the catalog illustrates a conversation. In this case, the order of each response is based on its allocation from a conversation or the list of strings. To instruct the data or corpus we have imparted the train function. In Doly, a lot of corpus data and effectiveness module that trained our chatbot in quickly for communication. We have to simply individualize the corpus data modules that we want to use. This training function approves the 'Doly' to be instructed using data from the dialog corpus that we made. While working with string type data in Python, it is practical to encounter errors. In this case, we use 'utf8' for encoding our corpus. When a chatbot trainer has provided with a data set, it builds the essential way in the chat bot's knowledge graph so that the questions or statement inputs and responses are correctly enacted. It sorted all the states from the corpus like graph sorted database in Fig 5.
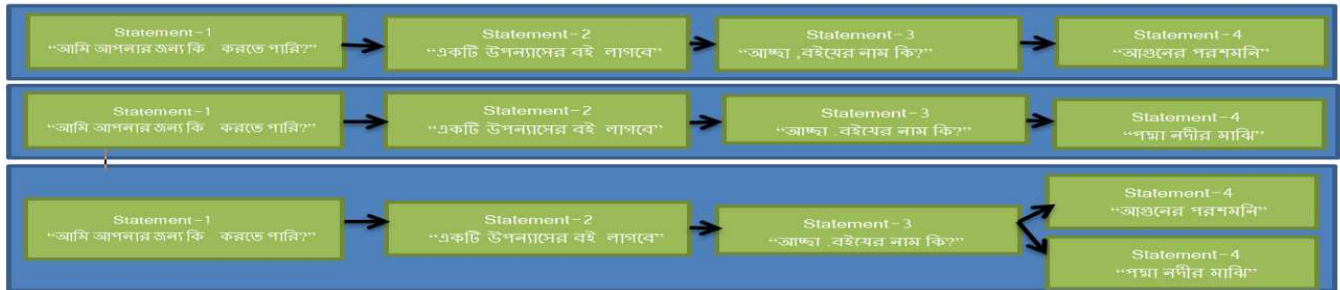

Fig. 5. Training data.

3) Input Adapters

Doly's input adapter is developed to authorize the chatbot to have a versatile method of adoption or restoring input from a given source. Input adapter is applied for getting input from some source, and then to convert it into a format that 'Doly' can access. So we created a format for the affirmation of our inputted data.
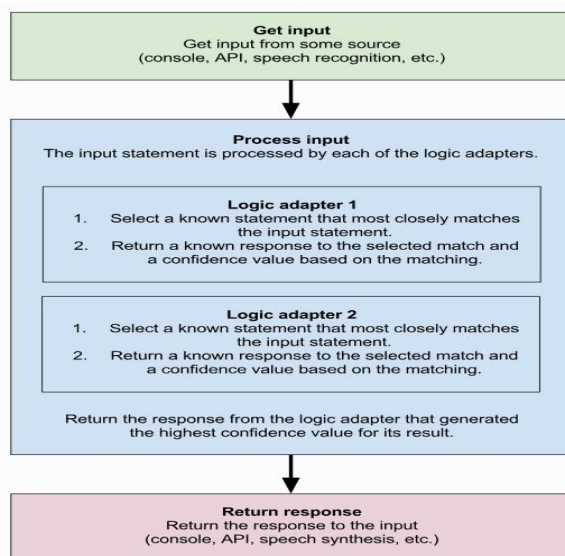

Fig. 6. Process flow diagram of Doly.

4) Process Unit

In order to process of adopted input the procedure function is needed. The input affirmations have been applied by an algorithm to find out the best likelihood valued reply for the query. The algorithm has been selected all the acquainted affirmations that most closely resemble with the input affirmation. It will return the acquainted repercussion to the selected match and a confidence estimate value based on matching after computation of each of the response. Here the confidence score is the likelihood value of the response. The algorithm will return the explanation that generated the highest likelihood value for itself.

5) Algorithams

To create develope 'Doly' we apply different types of machine learning methods to yield its responses. We also use machine learning algorithm like search algorithm for finding the appropriate list of matching results from the corpus and also use a Bayesian algorithm to choose the best possible answer form list of matching result.

In probability theory, Bayes' theorem is also known as Bayes' law or Bayes' rule which recounts the probability of a happening, based on previous knowledge of conditions that may be related to the happening. The Bayesian hypothesis is the most common application of Bayes' theorem that is a statistical hypothesis. Testing of the drug, productions, materials as well as computing the entire output of a company based on machines are the one kinds of example.

This theorem states mathematically in the following equation:

$$M(A \mid B) = (M(B \mid A) \ M(A))/(M(B))$$

where A and B are events and M(B) ≠0 and M(A│B) is a conditional probability of event A happening given that B is true. Similarly, M(B│A) is also a conditional probability happening given that A is true. Yet M(A) and M(B) are the probabilities of watching A and B independently of each other which is called the marginal probability. In a general way some of the various machine learning and artificial intelligence way that are implemented throughout Doly's methodology.

Search Algorithm: A searching system is a basic principle to find out data form given corpus or database in artificial intelligence. Search is the most momentous part of how Doly quickly and efficiently recuperate the possible statements that it can respond to the user. A few ways of characteristics that help the Doly select a response include I.The similarity of an input statement to known statements II.The frequency in which similar known responses occur. III.The likeliness of an input statement to fit into a category that known statements.

Classification Algorithm: In various types of logical adapters 'Doly' make sense to conduct Naive Bayesian classification algorithms to calculate if an input question is gathered to a particular set of criteria that take a response to be generated from that logic adapter. So simply we can mention that Doly is a Search-based artificial intelligence technique. For the sake of Search-based artificial intelligence, there are also some following optional algorithms implemented in Doly that can be used as the 'statement_comparison_function' for a logical adapter.

Levenshtein distance: Levenshtein distance or editing distance is dynamic programming algorithms where we can measure the minimum number of single characters are dissimilar between two words or strings. In Edit distance there are different sets of string operations such as removal, insertion, or substitution. The Edit distance follow some properties such as (a)A non-negative cost has been edited by every operation (b)For each operation, there is a reverse operation with equal costs.

There are several applications in Levenshtein Distance such as Natural Language Processing and Computational biology.

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & if \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} \end{cases} \quad or$$

Equation 1: Levenshtein Distance between two string

Here $a_i \ and \ b_j$ is an exponent. If $(a_i = b_j)$ means the contained character of ai and bj are not unequal there are unchanged and edit=0. Otherwise if $(a_{i!} = b_j)$ then edit=1. If we want to strings lemmatization than we have to check the entire corpus of root words with the inserted sentences through the help of minimal edit distance algorithm. The minimal edit can be greater than zero. If a corpus's word and an inserted string are equal then the edit will be zero or it is changeable.

Synset Distance: For calculating the synonymous similarity between statements, NLTK provides Synset distance. In Bengali words, we developed a corpus of synonymous similarity which covers to implement for data.

Jaccard Similarity: The Jaccard index also called the Jaccard similarity coefficient is a statistic for comparing the similarity and diversity of sample sets which are distant. The Jaccard similarity measured the similarity between finite sample sets and is defined as the ratio between the size of the intersection and union of sample sets. If P and Q be two sets the Jaccard index formula is given

$$J_{index}(P,Q) = \frac{|P \cap Q|}{|P \cup Q|} * 100 = \frac{|P \cap Q|}{|P|+|Q|-|P \cap Q|} \quad here,$$
$$0 \leq J_{index}(P,Q) \leq 1$$

The Jaccard distance measures dissimilarity between sample sets that are complementary to the Jaccard coefficient.

$$J_{distance}(P,Q) = 1 - J_{index}(P,Q)$$

Now, let an example using set notation and Venn-diagram

Fig-2: Jaccard similarity.

P= { দূরত্ব, বেগ, বল, চাপ, কাজ },Q= { কাজ, চাপ,গতি,স্বরণ}

$$J_{index}(P,Q) = \frac{|\{কাজ, চাপ\}|}{|\{দূরত্ব,বেগ, বল, চাপ, কাজ, গতি,স্বরণ \}|} = \frac{2}{7} = 0.28$$
$$And \ J_{distance} = 1 - J_{index}\% = 1 - 0.28 = 0.72$$

So, we conclude four postulates such as (a) If $J_{index} = 50$ similar which means the sets share all members.(b)If $J_{index} \ll 50$ similar which means the sets share few members. (c) If $J_{index} \cong 50$ similar which means the sets share half of the members. (d) If $J_{index} = 0$ similar which means the sets share no members.

D. Logic Adapters

The Logical adapters are identified as the logic for making responses of a conferred input statement or question by a Chatbot. The logical adapter which Doly used can be determined by applying the logic adapters parameter to the recruitment path of the logic adapter that needs to use. It is pretty possible to input any number of logic adapters for our chatbot in order to make jurisprudence. If multiple adapters are exercised, then the chatbot will retire the response with the best-enumerated confidence value.

E. Response Selection Method

Response selection methods determine which response should be used in the event that multiple responses are generated within a logic adapter. Doly uses Statement objects to hold information about things that can be said. An important part of how a chat bot selects a response is based on its ability to compare two statements to each

other. This module contains various text comparison algorithms designed to compare one statement to another. We use the get first response method for the selection of a response. This method takes the input statement and selects the statement in the knowledge base which closely matches the input to the chatbot from a list of statement options to choose a response from.
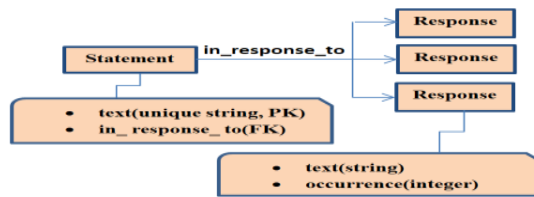

Fig. 7. Response System.

V. RESULTS & ANALYSIS

The difficulty of appraisal is intrinsic as each conversation having an effect on each other and the equivalent conversation will not happen to exceed once; one slightly different answer will lead to a completely different conversation [12]. So, for the excellent improvement of the chatbot, we compared our chatbot system with previous existing chatbots. When we checked the 2776 questions releted to inserted corpus on our chatbot system ,it returned 2448 correct reply.That means 88 % accurate.
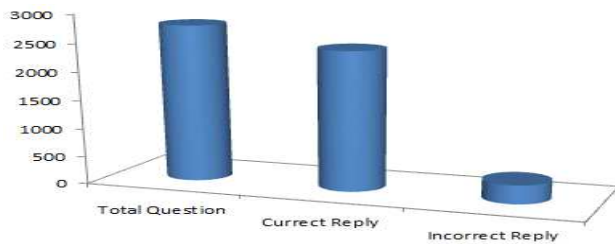

Fig. 8. Results Analysis.

VI. COMPARISON BETWEEN OTHERS CHATBOOK

Our purpose is creating Bengali chatbot in the sake of the educations system of Bangladesh. So, we compared 'Doly' with two chatbots which are Neural Conversational Machine (NCM) and Cleverbot. Neural Conversation Machine is a chatbot which can be developed with the end-to-end and this model chats by predicting the next sentence given the previous sentences or sentence with a few handcrafted rules [12]. So NCM is a generative based conversational agent whereas our task is a retrieval-based system. We give some random questions to the NCM and our system for the comparative study between the two. It responds to the input by finding how a human responded to that input when it was asked, in part or in full, by Cleverbot. In this case, 'Doly' acted pretty good and responded with the best way.

VII. FUTURE WORK

In future works, we plan to enable 'Doly' among all Bengali native speaker students as in the sake of effective educations tool with the Bengali language. We will create a chatbot with the help of neural network technique as well as a sequence to sequence chatbot via Bengali Language Corpus. Also, we will try to generate own large Bengali Language corpus and try to develop a great system for Bangla Natural Language Processing (BNLP).

VIII. CONCLUSION

This research is a pioneering work in the field of dialogue system in Bengali. The main challenge of this work is to create a chatbot based on the accurate knowledge base. Due to the lack of large dataset, we implemented a retrieval based closed domain chatbot which will converse with the user based on the pattern matching algorithm and will improve its performance measure by learning from the interaction. Our work will provide a Bengali conversation corpus which will assist in the construction of a system for Bengali Language Processing research.

IX. REFERENCES

[1]. Weizenbaum, J.: Contextual understanding by computers. Commun. ACM 10(8), 474–480 (1967)
[2]. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 36 –45 (1966)
[3]. ALICE:Artificial Intelligence Foundation. http://www.alicebot.org/bios/rechard wallace.html
[4]. Breazeal, C.: Designing Sociable Robots. MIT Press, Cambridge (2002)
[5]. Vinyals, O., Le, Q.: A neural conversational model. In: Proceedings of ICML Deep Learning Workshop (2015)
[6]. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput. Hum. Interact. (TOCHI) 12(2), 293–327 (2005)
[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. Al Mag. 22(4), 67–83 (2001)
[11]. Augello, A., Gentile, M., Weideveld, L., Dignum, F.: A model of a social chatbot. In: De Pietro, G., Gallo, L., Howlett, R.J., Jain, L.C. (eds.) Intelligent Interactive Multimedia Systems and Services 2016. SIST, vol. 55, pp. 637–647. Springer, Cham (2016). doi:10.1007/978-3-319-39345-2_57
[12]. Hill, J., Ford, W.R., Farreras, I.G.: Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations. Comput. Hum. Behav. 49, 245–250 (2015)
[13]. Pan, X., Slater, M.: A preliminary study of shy males interacting with a virtual female. In: Presence 2007: The 10th Annual International Workshop on Presence (2007)
[14]. Crutzen, R., Peters, G.-J.Y., Dias Portugal, S., Fisser, E.M., Grolleman, J.J.: An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J. Adolesc. Health 48, 514–519 (2011)
[15]. Shawar, Bayan Abu, and Eric Atwell. *A comparison between Alice and Elizabeth chatbot systems*. University of Leeds, School of Computing research report 2002.19, 2002.
[16]. Abdul-Kader, Sameera A., and J. C. Woods. "Survey on chatbot design techniques in speech conversation systems." *International Journal of Advanced Computer Science and Applications* 6.7 (2015).
[17]. Shrestha, Lokesh, and Kathleen McKeown. "Detection of question-answer pairs in email conversations." *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, 2004.
[18]. Zhou, Liang, and Eduard Hovy. "Digesting virtual geek culture: The summarization of technical internet relay chats." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005.