

Shiksha Mitra: An Assamese Language AI Chatbot Using Deep Learning

Surajit Sarma*, Dr. Nabankur Pathak

Department of Computer Science, Krishna Kanta Handiqui State Open University, Guwahati, Assam, India

ARTICLE INFO

Article History:

Accepted: 15 Oct 2023

Published: 12 Nov 2023

Publication Issue

Volume 9, Issue 6

November-December-2023

Page Number

48-57

ABSTRACT

This research paper presents “Shiksha Mitra”, an artificial intelligence chatbot that answers user queries in Assamese for educational purposes. The chatbot uses Assamese Natural Language Processing (ANLP) and deep learning techniques to identify relevant sentences and provide responses. Unlike many organizations that use English chatbots, this research aims to develop a data-driven, retrieval-based, closed domain chatbot that can interact with users in Assamese. The chatbot is trained with corpus data encoded in UTF-8 format using a train function adapter. A feedforward neural network is used to find the best match from the corpus and generate a suitable answer.

Keywords: Artificial Intelligent, Assamese Corpus, Chatbot, Deep Learning, Natural Language Processing, Feed Forward Neural Network,

I. INTRODUCTION

Educational institutions are not just places to learn anymore. They are also businesses that offer many services to their students. These services help students with their studies, careers, mental health, money, and other needs. The aim is to give students a complete support system that helps them overcome the difficulties of higher education and achieve their goals.

To do this, educational institutions need to be available 24 hours a day to answer any questions or problems that students have. This can be anything from simple queries about campus life to urgent help in case of an accident or illness.

Institutions can choose to use either a chatbot or a call center to provide this support. Chatbots are becoming more popular because they can quickly respond to many types of student questions. They are especially helpful for support after normal hours. On the other hand, call centers can give a more personal touch, with trained staff who can help students with more complicated issues.

The change to service-based education shows that educational institutions care about giving students a well-rounded education that covers both academic and non-academic aspects. By offering 24-hour support services, educational institutions can make sure that their students have everything they need to succeed in and out of the classroom.

A Chabot is a computer program that provides a platform for user to do human conversation in its natural form. Chatbot is a Software program that generates response based on given input to mimic human conversations in text or voice mode (Jia et al., 2003). Chatbot and similar applications are programmed with predefined language scripts in order to formulate meaningful responses to specific enquiries (Gkoutziouli et al., 2022). "Chatbots are intelligent conversational computer systems designed to mimic human conversation to enable automated online guidance and support" (Caldarini et al., 2022). Nowadays, Chatbot are being used in different applications and fields in the areas of education, health, e-commerce, entertainment etc. Chatbot can be a useful tool to support services in different fields to the user (T et al., 2018). Chatbot is a multiple user system that can provide support and assistant to multiple users at a time (Costa et al., 2018). "Human-Computer Interaction that characterizes dialogue between human and computer is gaining momentum in computer interaction techniques. This type of program is called a Chatbot or chatterbot." (Madhumitha et al., 2019).

In today's rapidly advancing world, technology has become an indispensable tool for progress and development. In India, digital interventions have brought about a revolution in even the most traditional government services, paving the way for incredible progress. As the country marches forward towards a brighter future, the logical next step is to introduce multilingual, knowledge-rich, and intelligent chatbot. These digital assistants have the potential to revolutionize the way government services are delivered to citizens.

The introduction of multilingual, knowledge-rich, and intelligent chatbot in government services marks a significant milestone in India's journey towards progress and development. By leveraging the power of technology, the country can create a more efficient and responsive government that delivers personalized services to its citizens, paving the way for a brighter and more prosperous future.

The rest of this paper is organized as follows. Section II reviews the related literature. Section III describes the methodology used. Section IV presents and discusses the results. Section V concludes the paper.

II. LITERATURE REVIEW

In the year 2003, Poongkuzhali, an Artificial Intelligence-based Tamil chatbot was developed. The chatbot responds to the user's query based on the input. In this case, the concept of a pre-existing conversation topic is employed. The ability to read user input in any form, regardless of grammatical format, has been specified by a set of decomposition rules. The response is generated using the knowledge base's collection of reassembly rules. The response is rephrased to match the user's query. In order to manage references to prior inquiries, history data is kept. The chatbot is of order n in terms of time complexity, where n is the number of terms in the input (Kalaiyarasi et al., 2003).

In the year 2014, UMAIR, an Urdu-based customer service Chabot was created. To deal with the issues that occur in Urdu languages, this Chatbot uses scripting languages, the Word Order Wizard string similarity method, and a novel engine as its design. The controller is the Chabot's fundamental core, and it is in charge of all communication. Conversation Manager (CM) and Path Manager (PM) are the two sorts of managers in the Chatbot (PM). CM guarantees that the user stays focused on the topic's goal. The decision tree in the architecture is loaded by PM so that the debate advances to the desired leaf node. Responses are provided by scripting languages in the form of images, maps, and forms. An utterance cleaner is used to remove the special character and normalize the user input. The accuracy rate predicted is 83.7 percent (Kaleem et al., 2014).

In the year 2017, Golopo, a Bengali chatbot, was created. Golpo is a data-driven, retrieval-based closed domain chatbot that can communicate in Bengali. The bot is mostly built on a pattern matching algorithm that will learn from encounters

to enhance its efficiency. A Bangla corpus that was personally developed is used to chat in basic Bangla. It responds to user enquiries quickly and efficiently using a chat interface based on its knowledge base. This chatbot's backbone is a language-independent NLP library with a learning mechanism. The input is first matched with database entries, then a confidence score for the matched sentences is produced, and lastly the one with the greatest confidence score is chosen as the answer to the input. The Best Match Adapter is used to provide a response that is based on the best-known match to a given statement. The Jaccard Similarity function then compares the known and input statements using a Jaccard Index. This index is maintained at 50%. When compared to the Neural Conversational Model (NCM), a generative-based model, the Chatbot exceeds the current state-of-the-art model (Orin et al., 2017).

In the year 2018, AskDISHA, a multilingual chatbot that can communicate in both English and Hindi was developed. IRCTC and CoRover collaborated to create an AI-powered chatbot. Artificial intelligence, machine learning, and natural language processing are all used by AksDISHA. This Chatbot was created in order to respond to questions from users who are having issues with Indian Railway ticket reservations. This Chatbot provides an immediate response to questions on ticket reservations, refund status, cancellation, reservation, PNR search, and other IRCTC travel packages and offerings. The accuracy rate of answering passenger queries is 90% (Dey et al., 2020).

In the year 2019, AMBER, a Chabot that is for the Marathi language was developed. AMBER stands for "A Marathi Based Eliza Chatterbot." Sentence Generation is used by this Chatbot. For semantic determination, libraries such as the paraphrasing detection and a variety of other frameworks were used. The alignment level in idioms and phrasal chunks is calculated using the Word Alignment algorithm. Natural Language Processing activities such as tokenization, part of speech, and

transliteration are all affected by AMBER. To locate suffixes, a phonetic and stripping technique is utilized (Shah et al., 2017).

In the year 2019, Doly, another Bengali chatbot, was created. Doly is a Bengali Chatbot that responds to user queries in Bengali. It was built with the intention of being used in the educational sector. The pillars of this AI-powered chatbot are Machine Learning (ML) methods and Bengali Natural Language Processing (BNLP). To explore the corpus for similar findings, the ML method works similarly to a search algorithm. Then, using Nave Bayesian, the suitable response is generated from the list. The dialogue corpus for the chatbot is derived from a simple Bengali Language corpus and saved in.yml format. The entire data-set is pre-processed in order to train the prepared data-set. Anaphora, stop word removal, and lemmatization are some of the pre-processing techniques used. Hobbs' Anaphora Resolution algorithm is used by the bot (Kowsher et al., 2019).

In the year 2019, Disha, a Bangla chatbot that specializes in health care was developed. The Chabot employs machine learning to facilitate human-to-human communication with the virtual medical assisant. This Chabot converses via text with the help of a knowledge base. This Chabot diagnoses the diseases also keeps records of the patient's health and advises the patient against health hazards based on the symptoms. This chatbot interacts with the user by inquiring about the user's age, name, blood group, and other basic information. In Bangla, a specific dataset for disease classification has been created. The diseases' names, as well as the type of doctor, are mapped in the dataset. Once the ailment has been recognized, the chatbot will make recommendations for the type of doctor that is needed. SVM was chosen for classification and accuracy. This Chabot has a 98.89 percent accuracy rate (Rahman et al., 2019).

In the year 2020, Chatterbot Python Library based Malayalam Chabot was developed. Using the Chatterbot architecture, this Chatbot was created for the Malayalam language. It makes advantage of

Python's NLP and Machine Learning packages. The database records are compared to the queries in this Chatbot, and a confidence value is calculated. This context is processed whenever a user enters a sentence in the web interface, and the known responses based on the context are obtained. The user is shown the response with the highest confidence score among the selected responses. The corpus is YAML-formatted and extracted from the Malayalam Tourism corpus. For finding the closest match, the Best Match logic adapter is employed. The precise solution is generated by Jaccard indices. In order to engage with the user, Levenshtein is used (Sandhini et al., 2018).

In the year 2021, Nubot, a Chatbot that specializes in Roman Urdu was developed. For unstructured Roman Urdu language, this Chatbot delivers proper responses. This Chatbot was built using the RASA framework. The purpose is classified using the RASA NLU module-based corpus. A knowledge graph is used with RASA Framework to maintain track of the discussion history for a domain specific natural language paradigm. The overall response generating accuracy percentage is 82.1 percent (Shabbir et al., 2018).

Some research papers on IR systems for the Assamese language, which has few NLP tools found has been discussed below. One paper used the Vector Space Model and Assamese WordNet to make search results better (Sarmah et.al, 2013). Another paper used a logical way to answer questions in Assamese (Chakraborty et.al, 2015). This paper helped AI and NLP for the Assamese language and other languages and cultures. A third paper (Bhuyan et.al, 2019) used Assamese WordNet and Wikipedia to get information from a big collection of texts. The system worked well, but had some problems with some queries that Wikipedia did not have. The paper said some ways to make the system better, like using more texts, more web sources, and a better ranking way like the Vector Space Model. These papers show the progress and possibility of IR systems for the Assamese language,

which can help the people who use and study this language.

No AI chatbot based on deep learning that is specifically trained in the Indo-Aryan language of Assamese has been discovered as of the writing of this research paper.

III. METHODOLOGY

"Shiksha Mitra" is an AI Chatbot designed specifically for the Assamese language. It has been created using Assamese Natural Language Processing (ANLP) and deep learning techniques. The chatbot has been built on the Python programming language, using the Pytorch library. It uses a Feed Forward Neural Net, consisting of four hidden layers, to process user inputs and generate responses.

ANLP is a self-developed library that comprises a variety of functions designed to preprocess user requests and responses. It helps in converting the input into vector form, which can be easily understood by the Feed Forward Neural Net. The flow chart of the proposed system can be seen in "Figure 1".

The ANLP library pre-processes the user input and converts it into a vector. The vector form of the user's input is then passed on to the Feed Forward Neural Net. The neural network processes the input vector and generates a response with a high probability of being the correct output for the user's request. The chatbot uses deep learning techniques to improve its accuracy over time and to provide better responses to the user's queries. Overall, "Shiksha Mitra" is a sophisticated chatbot that leverages ANLP and deep learning techniques to provide a seamless user experience in the Assamese language.

Pre-processing

To prepare the training data for the chatbot, a series of methods were employed to pre-process the entire dataset. These methods included Assamese Word

Tokenization, Assamese Stop Word Removal, and Assamese Word Stemming.

Assamese Word Tokenization is a process that effectively splits a given string into meaningful units, or tokens. This process is crucial in ensuring that the chatbot can understand and process individual words in user queries.

data, allowing the chatbot to focus on the most relevant and significant words.

Assamese Word Stemming is a process that generates the root form of words, a crucial step in ensuring that the chatbot can understand and process the meaning of words. This process involves removing prefixes and suffixes to reveal the base word.

By employing these various methods for pre-processing the dataset, we were able to prepare an effective training dataset that enabled the chatbot to accurately understand and respond to user queries in the Assamese language.

Features Extraction Process

In order to train our deep learning model, a series of preparatory steps were undertaken. Firstly, all the sentences in the pattern had to be converted into vectors. To achieve this, a function called "bag of words" was created, which converted all the sentences in the pattern into an array of vectors. This array was then used as input for the feed forward neural network, which was trained to produce different probabilities for different tag levels or classes.

To prepare the input data for the neural network, the sentences in the pattern were broken down into arrays of single words, as depicted in " Figure 2", using a process known as tokenization. An array was created for each pattern, with the size of each array being equal to the number of single words in the sentence. For each position in the array, a value of 1 was placed if the word in the pattern matched the corresponding word in the array of single words, and 0 otherwise.

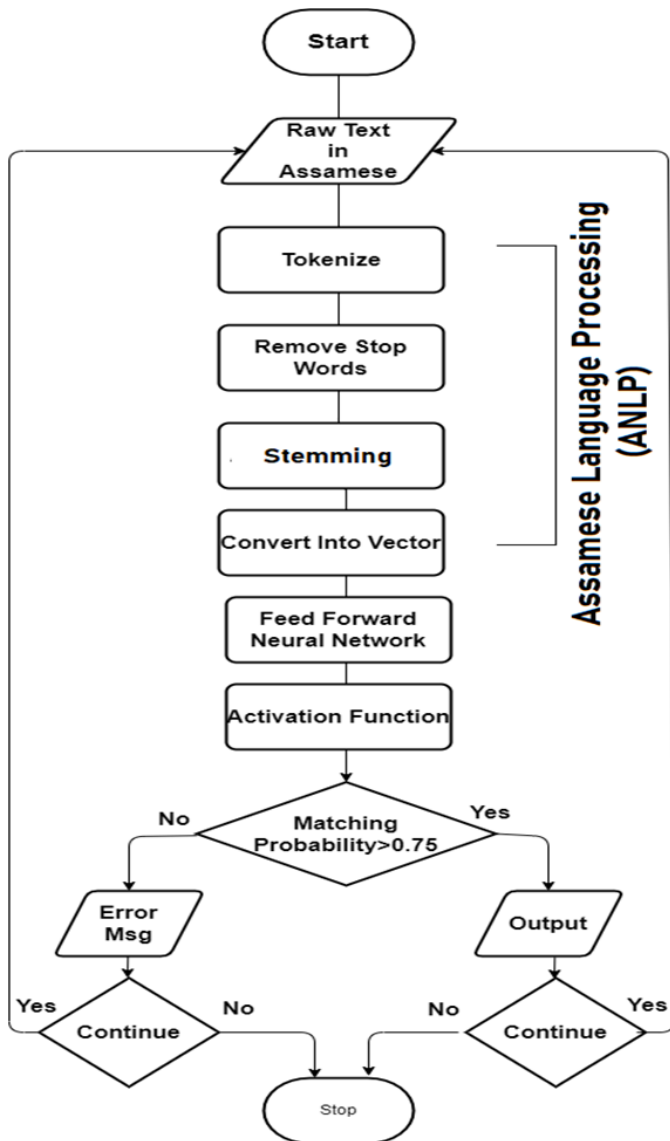


Fig. 1. Flowchart of Proposed system

Assamese Stop Word Removal, on the other hand, is a feature that removes all stop words from the list of tokenized words. This step is important in eliminating redundant and unnecessary words from the training

The tag level, which ranged from 0 to n, was also assigned to each pattern. The bag of words array acted as the X vector, while the tag level acted as the Y vector, for the model to be trained. By undertaking these preparatory steps, we were able to convert the pattern sentences into a format that could be used to train the feed forward neural network.

Single Array of Words								Y		
	অনলাইন	মতুল	পরিশেষ	পোস্টে	মোড	নামভর্তি	পঠ্যক্রম		তালিকা	আবহু
অনলাইন মতুল পরিশেষ	1	1	1	0	0	0	0	0	0	0 (মতুল)
পোস্টে মতুল অনলাইন মোড	1	1	0	1	1	0	0	0	0	
নামভর্তি পঠ্যক্রম	0	0	0	0	0	1	1	0	0	1 (নামভর্তি)
নামভর্তি তালিকা আবহু	0	0	0	0	0	1	0	1	1	
	X									

Fig. 2. Single array of words (Bag of words)

Feed Forward Neural Network Model (FFNN)

One of the main advantages of using a feed forward neural network is its ability to process large amounts of data in a relatively short amount of time. The feed forward architecture is designed to process input data

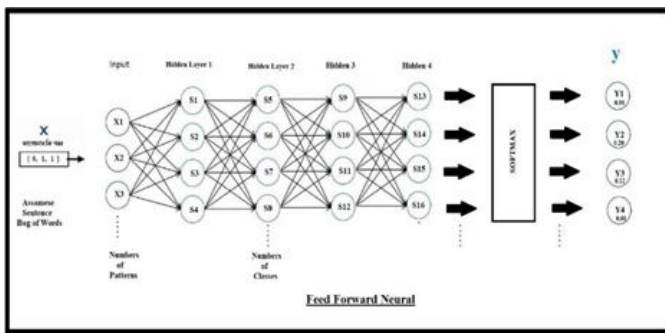


Fig. 3. Format Feed forward neural network

in a single direction, from input to output layer, without any loops or cycles. This allows for efficient processing of data in a fast and accurate manner. Another advantage of using a feed forward neural network is its ability to handle non-linear data. Non-linear data can be difficult to model using traditional statistical methods, but a feed forward neural network is capable of learning and representing non-linear relationships between inputs and outputs. This makes it a powerful tool for tasks such as image and speech recognition, natural language processing, and predictive modeling.

Additionally, the four hidden layers used in the "Shiksha Mitra" chatbot allows for the network to learn and extract more complex features from the input data. This can lead to higher accuracy and better performance when compared to a network with fewer hidden layers. Overall, the use of a feed forward neural network with multiple hidden layers provides a powerful and efficient tool for building complex models for a variety of applications. This Feed forward neural network is illustrated in "Figure 3".

Description of Calculation of neuron (Figure 4) in single unite of proposed feed forward neural network:

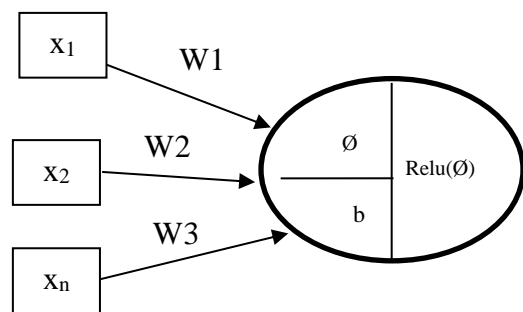


Fig.4. Structure of Neurons

1. Neurons: A neuron, also known as a node or perceptron, is a basic unit of a feedforward neural network.
2. Inputs: Neurons receives inputs from other neurons or directly from the input layer, and applies a set of weights to the inputs. Let us consider the input as given below:

$$x_1, x_2 \dots x_n$$

3. Weights: Each connection between neurons in adjacent layers has a weight associated with it. These weights determine the strength of the connection and are adjusted during the training process to minimize the error. Let us consider the weight as given below:

$$w_1, w_2 \dots w_n$$

4. Bias: Each neuron has a bias term added to its weighted sum of inputs. The bias term helps to shift the activation function left or right and can be adjusted during the training process. Let us consider the biases as given below:

$$b_1$$

5. Activation Function: The activation function is applied to the weighted sum of inputs and the bias to produce the output of the neuron. We have used ReLU activation functions in proposed feed forward neural network

$$f(\phi) = \max(0, \phi) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } x > 0 \end{cases}$$

6. Output: Neurons adds a bias term, and passes the result through an activation function to produce an output, which is then passed to the next layer of neurons.

This is representing mathematically as follows:

$$\begin{aligned} \phi &= w_1x_1 + w_2x_2 \dots w_nx_n + b_1 \\ &= \sum_{i=2}^n (w_ix_i) + b_1 \end{aligned}$$

Loss Function

The loss function plays a crucial role in guiding the learning process by indicating the necessary corrections to be made in the model. To compare the predicted probability distribution with the true probability distribution, we will use the cross-entropy loss. In the case of binary classification, the cross-entropy loss is defined as follows:

$$\begin{aligned} \text{Log Loss}(L) &= -\frac{1}{N} \sum_{i=1}^n y_i \log(f(y_i)) \\ &\quad + (1 - y_i) \log(1 - f(y_i)) \end{aligned}$$

Dataset

Our team has taken great strides to develop a dialogue corpus that caters to the Assamese language. We meticulously curated a few uncomplicated yet effective corpora that were saved in json format for easy accessibility. Through our rigorous efforts, we have managed to accumulate an impressive 2261 questions along with their corresponding answers.

The structure of this meticulously crafted corpus can be seen in” Figure 5”. Our approach involved incorporating a wide range of topics that were relevant and relatable to the Assamese-speaking community. We wanted to ensure that our corpus was comprehensive enough to cater to a broad range of users and their varying needs.

```
{ "tag": "পৰামৰ্শ",
  "patterns": [
    "শিক্ষাৰ্থীসকলৰ বাবে পৰামৰ্শ দিয়াৰ উদ্দেশ্য কি?",
    "শিক্ষাৰ্থীসকলে পৰামৰ্শৰ বাবে কাক যোগাযোগ কৰিব লাগে?",
    "অনলাইন পৰামৰ্শৰ প্লাছড"
  ],
  "responses": [
    "শিক্ষাৰ্থীসকলৰ বাবে কাউন্সেলিং অনুষ্ঠিত হ'ব যিবোৰ নিজ নিজ অধ্যয়ন কেন্দ্ৰৰ বাবে ভেঙেলোকে কৰিব যোগাযোগ কৰিব লাগিব নিজ নিজ অধ্যয়ন কেন্দ্ৰ সহ-coordinators. ইয়াৰ উপৰিও KKHSSOU চহৰ অধ্যয়ন কেন্দ্ৰই অনলাইন পৰামৰ্শৰ আয়োজন কৰে সমগ্ৰ ৰাজ্যখনৰ সকলো শিক্ষাৰ্থীৰ বাবে শ্ৰেণীসমূহ যিবোৰ অধ্যয়ন কেন্দ্ৰত ভেঙেলোকে...নামভৰ্তি কৰা হৈছে",
    "শিক্ষাৰ্থীসকলৰ বাবে কাউন্সেলিং অনুষ্ঠিত হ'ব যিবোৰ নিজ নিজ অধ্যয়ন কেন্দ্ৰৰ বাবে ভেঙেলোকে কৰিব যোগাযোগ কৰিব লাগিব নিজ নিজ অধ্যয়ন কেন্দ্ৰ সহ-coordinators. ইয়াৰ উপৰিও KKHSSOU চহৰ অধ্যয়ন কেন্দ্ৰই অনলাইন পৰামৰ্শৰ আয়োজন কৰে সমগ্ৰ ৰাজ্যখনৰ সকলো শিক্ষাৰ্থীৰ বাবে শ্ৰেণীসমূহ যিবোৰ অধ্যয়ন কেন্দ্ৰত ভেঙেলোকে...নামভৰ্তি কৰা হৈছে",
    "শিক্ষাৰ্থীসকলৰ বাবে কাউন্সেলিং অনুষ্ঠিত হ'ব যিবোৰ নিজ নিজ অধ্যয়ন কেন্দ্ৰৰ বাবে ভেঙেলোকে কৰিব যোগাযোগ কৰিব লাগিব নিজ নিজ অধ্যয়ন কেন্দ্ৰ সহ-coordinators. ইয়াৰ উপৰিও KKHSSOU চহৰ অধ্যয়ন কেন্দ্ৰই অনলাইন পৰামৰ্শৰ আয়োজন কৰে সমগ্ৰ ৰাজ্যখনৰ সকলো শিক্ষাৰ্থীৰ বাবে শ্ৰেণীসমূহ যিবোৰ অধ্যয়ন কেন্দ্ৰত ভেঙেলোকে...নামভৰ্তি কৰা হৈছে"
  ]
}
```

Fig. 5. Format of dataset corpus

Assamese natural language processing (ANLP) library

The Assamese language has a unique font that poses a challenge for processing. We used "utf8" encoding to make the system read and process the language effectively. Since there is no public Assamese NLP library, we created an NLP library package for it.

Our NLP package has several features for processing the Assamese language. First, it has an Assamese language detection function that identifies when a statement uses the language.

Second, it has an Assamese sentence tokenization function that splits a paragraph into sentences. It also

has an Assamese word tokenization function that splits a sentence into words, and an Assamese letter tokenization function that splits a word into characters.

Third, it has a fuzzy sentence similarity checking function that determines if two sentences are similar. This helps the chatbot give accurate and relevant responses to user queries.

Fourth, it has an Assamese word stemming function that finds the root word of a word in the WordNet database. Lastly, it has an Assamese stop word removal feature that removes stop words from sentences or words.

These features make our NLP library package comprehensive and efficient. It helps us develop a chatbot system that can process and respond to Assamese language queries effectively.

IV.RESULTS AND DISCUSSION

We have used the scikit-learn Python package to test the model accuracy and performance. We have divided this section into two parts: (A) Accuracy analysis and (B) Performance analysis.

A. Accuracy analysis

We have tested the accuracy of our proposed model using inbuilt function of scikit-learn. The result has been presented in the Table I.

TABLE I
ACCURACY COMPARISON OF MODELS

Models	Accuracy (%)
FFNN	89.11
SVM	88.74
Naive Bayes	84.11

The table shows the accuracy of three machine learning models: FFNN, SVM, and Naive Bayes, on Assamese text classification task. Based on the table, we can see that FNN has the highest accuracy (89.11%) among the three models, followed by SVM (88.74%) and Naive

Bayes (84.11%). This means that FFNN is the best model for this classification task on our self-made dataset, as it makes the most correct predictions out of all the models.

B. Performance analysis.

Accuracy alone may not be enough to evaluate the performance of a model, especially when the data is imbalanced or there are multiple classes. In such cases, other metrics such as precision, recall, F1 score may be more informative and useful. So we have used the inbuilt function of scikit-learn to evaluate the performance based on metrics: precision, recall and F1 score. The result has been presented in the Table II.

TABLE III
PERFORMANCE COMPARISON OF MODELS

Models	Precision		Recall		F1 Score	
	Macro	Weighted	Macro	Weighted	Macro	Weighted
FFNN	0.76	0.89	0.64	0.89	0.68	0.88
SVM	0.7	0.9	0.61	0.89	0.63	0.89
Naive Bayes	0.39	0.79	0.27	0.84	0.29	0.8

The table shows the performance of three machine learning models: FFNN, SVM, and Naive Bayes, on Assamese classification task on our self-made dataset. The performance is measured by three metrics: precision, recall, and F1 score. These metrics are calculated in two ways: macro and weighted.

Based on the table, we can make some observations:

FFNN has the highest macro precision, recall, and F1 score among the three models. This means that FFNN is better at identifying and finding all classes equally, without favouring any particular class on our dataset. SVM has the highest weighted precision, recall, and F1 score among the three models. This means that SVM is

better at identifying and finding the most frequent classes, while ignoring the less frequent ones. Naive Bayes has the lowest macro and weighted precision, recall, and F1 score among the three models. This means that Naive Bayes is worse at identifying and finding any class, regardless of their frequency.

V. CONCLUSION

This study is a significant achievement in the research on Assamese AI chatbot. One of the main challenges was to build an accurate knowledge base for the chatbot. This was hard because there were no resources or public datasets available, so we had to make our own corpus. We developed a chatbot model using feedforward neural network and trained it with our dataset. We compared our FFNN model with traditional machine learning models and achieved better accuracy than them. This research has enabled us to develop a dialogue system for Assamese Language Processing, which is an important step for improving natural language processing in the Assamese language. In the future, we plan to develop a chatbot that can chat in Assamese. We will use different deep learning models and new technologies to build this chatbot. We will also work on enhancing the Assamese corpus and adding more features to our Assamese Natural Language Processing Library (ANLP).

VI. REFERENCES

- [1] Barman, A. K., Sarmah, J., & Sarma, S. K. (2013). WordNet based information retrieval system for Assamese. In 2013 UKSim 15th International Conference on Computer Modelling and Simulation (pp. 10-12). IEEE. <https://doi.org/10.1109/UKSim.2013.90>
- [2] Bhuyan, M. P., Purkayastha, R., Sarma, S. K., Sarmah, S., Sarma, P., & Deka, V. (2019). Information retrieval in Assamese using Wordnet & Assamese Wikipedia. International Journal of Scientific & Technology Research, 8(9), 01-10.
- [3] Caldarini, G., Jaf, S., & McGarry, K. (2022). A literature survey of recent advances in chatbots. MDPI. Multidisciplinary Digital Publishing Institute. Retrieved October 3, 2022,
- [4] Chakraborty, R., & Sarma, S. (2015). Modelling a question-answering system using structured representation of Assamese text. International Journal of Artificial Intelligence & Applications, 6, 01-10. <https://doi.org/10.5121/ijaia.2015.6601>
- [5] Costa, P. (2018). Conversing with personal digital assistants: On gender and artificial intelligence. Journal of Science and Technology of the Arts, 10(6), 59-72.
- [6] Dey, S., & Shukla, D. (2020). Analytical study on use of AI techniques in tourism sector for smarter customer experience management. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-6). IEEE. <https://doi.org/10.1109/iccsea49143.2020.9132925>
- [7] Gkoutziouli, K. (2022). Untraining bias: AI chatbots and language. Academia.edu . Retrieved October 22, 2022
- [8] Jia, J. (2003). The study of the application of a keywords-based chatbot system on the teaching of foreign languages. arXiv.org . Retrieved October 29, 2022
- [9] Kalaiyarasi, T., & Parthasarathi, R. (2003). Poongkuzhali-an intelligent Tamil chatterbot. Semantic Scholar . Retrieved November 5, 2022, from 4
- [10] Kaleem, M., O'Shea, J., & Crockett, K. A. (2014). Development of Umair the Urdu conversational agent for customer service. Semantic Scholar . Retrieved November 7, 2022.
- [11] Arafat Khan and Nayeem Al Noman (2019), "Doly: Bengali chatbot for Bengali education", In 2019 1st International Conference on Advances in Science, Engineering and Robotics

Technology (ICASERT) (pp. 1-6).IEEE.
<https://doi.org/10.1109/icasert.2019.8934592>

- [12] Madhumitha, S., Keerthana, B., & Hemalatha, B. (2019). Interactive chatbot using AIML. Ijana , 1(1), 62-66.
- [13] Orin, T. D. (2017). Implementation of a Bangla chatbot. BRAC University . Retrieved November 21, 2022
- [14] Rahman, M. M., Rahman, M. A., Islam, M. S., & Hossain, M. A. (2019). Disha: An implementation of machine learning based Bangla healthcare chatbot. In 2019 22nd International Conference on Computer and Information Technology (ICIT) (pp. 1-6). IEEE. <https://doi.org/10.1109/iccit48885.2019.9038579>
- [15] Sandhini, S., Sreeja, P., & Sreeja, P. (2018). A proposal of chatbot for Malayalam. International Journal of Computer Sciences and Engineering , 6(6), 21-25. <https://doi.org/10.26438/ijcse/v6si6.2125>
- [16] Shabbir, J., Arshad, M. U., & Shahzad, W. (2021). NUBOT: Embedded knowledge graph with RASA framework for generating semantic intents responses in Roman Urdu. Semantic Scholar . Retrieved from 6
- [17] Shah Manthan, Jigneshkumar, Arnav Mediratta, Akshay Kundale, & Shubham Nangare (2017). AMBER chatbot and detection of paraphrases for Devnagari. Vishwakarma Journal of Engineering Research , 1(1), 1-5.

Cite this article as :

Surajit Sarma, Dr. Nabankur Pathak, "Shiksha Mitra: An Assamese Language AI Chatbot Using Deep Learning ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 6, pp.48-57, November-December-2023. Available at doi : <https://doi.org/10.32628/CSEIT2390572>
Journal URL : <https://ijsrcseit.com/CSEIT2390572>