BotSom - Somaiya Chatbot with NLTK

SLP IA2 Report

Group Members:
Shubh Gangar: 1911013
Kashish Jaisinghani: 1911016
Harshit Jain: 1911017
Devansh Shah: 1911052

Problem Definition:

A chatbot that can assist KJ Somaiya College students and faculty in accessing information and resources regarding academic programmes, fee structure, campus services, and other pertinent topics. It will be able to respond to frequent questions and direct users to the appropriate departments or resources. It will be incorporated with the university's existing information systems and will be accessible via multiple channels, including the university's website, social media platforms, and messaging applications.

Literature Survey:

Title	Description
Creating a ChatBot using the basic ML Algorithms	This article discusses chatbots founded on machine learning. The chatbot responds to weather questions with pre programmed responses. The article describes the cleansing and preprocessing of the dataset, the production of a bag-of-words representation of phrases, and the training of logistic regression, SVMs, and decision trees. Using accuracy and confusion matrices, this study examines each model. The article is concluded by chatbot advancements including intents and web interfaces. The article focuses on analysing and comparing models to enhance performance and presents fundamental ML approaches for chatbot construction.

Using NLP to Improve Chatbot Performance: Lessons from Industry Leaders	This article offers a detailed walkthrough on how to build a chatbot using natural language processing (NLP). It discusses the fundamentals of natural language processing (NLP), how to choose a platform on which to construct your chatbot, and best practices for the design of an efficient conversational user interface.
Using NLP to Improve Chatbot Performance: Lessons from Industry Leaders	This article investigates how industry leaders are making use of natural language processing (NLP) to enhance the functionality of their chatbots. It discusses subjects such as entity identification, sentiment analysis, and intent classification, and it presents instances of firms that are adopting these approaches to provide more successful chatbot experiences to their customers.
Building a Chatbot with Natural Language Processing	This article presents a step-by-step method to construct a chatbot using natural language processing (NLP). It teaches you the fundamentals of natural language processing, how to gather and preprocess data, as well as how to train and deploy your chatbot. Additionally, it provides code snippets that can assist you in getting started.
Ticketing Chatbot Service using Serverless NLP Technology	Using a user study, the "Ticketing Chatbot Service Using Serverless NLP Technology" paper evaluated the efficiency and efficacy of the chatbot service. The study determined that the chatbot was effective at comprehending user requests and providing appropriate responses, resulting in shorter completion periods and high user satisfaction. The conclusion of the paper is that serverless NLP technology is an efficient and scalable solution for developing chatbot services across multiple domains.
Investigating the user experience of customer service chatbot interaction: a framework for qualitative analysis of chatbot dialogues	The article proposes a methodology for qualitative analysis of chatbot dialogues in customer service contacts. The framework investigates chatbot user experience for customer service. A customer service chatbot was used to evaluate sixteen individuals. The framework uncovered user experience-related topics in chatbot conversations. The survey revealed that

	customers were pleased with the performance of the chatbot but desired more personalised responses and improved conversation quality. The authors contend that the framework can aid in the analysis of chatbot conversations and the comprehension of customer service user experience.
Designing a Chatbot Using Natural Language Processing	This article offers a detailed instruction on how to construct a chatbot using natural language processing (NLP). It teaches you the fundamentals of natural language processing, how to gather and preprocess data, and how to make use of machine learning techniques in order to construct and educate your chatbot. In addition to that, it has an in-depth walkthrough with examples of code at each stage.

Dataset:

The custom dataset was created by manual for a chatbot that can assist users with Somaiya Institute-related frequently inquired queries. The dataset contains 17 categories, including salutation, name, goodbye, address, course, information, library, fees, hours, canteen, scholarships, projects, admission, hostel, events, placements, and councils.

Each tag includes approximately 15 prompts and three responses. The chatbot chooses responses at random to provide a more natural conversation flow. In addition, the responses contain text output and relevant links to the Somaiya website, where users can find additional information.

Preprocessing:

The dataset underwent preprocessing of the text prompts using various techniques offered by the Natural Language Toolkit (nltk) library. The initial stage involved the removal of unnecessary characters or symbols that might hinder the training process of the model. The sentences underwent tokenization, a procedure that involves dividing them into distinct parts or tokens, typically consisting of individual words. Subsequently,

the tokens underwent a conversion process to lowercase in order to maintain uniformity and prevent complications arising from case sensitivity.

Following that, the tokens were stripped of their punctuation in order to concentrate exclusively on the semantic content of the words. The text that was cleaned later underwent transformation through the application of stemming and lemmatization techniques.

Stemming is a linguistic technique that involves the truncation of affixes from words to obtain their base form. This process is useful in minimising the count of distinct words in a given corpus. The PorterStemmer algorithm, a frequently employed stemming algorithm in the field of natural language processing, was utilised.

Lemmatization is a linguistic technique that bears resemblance to stemming. However, it entails the reduction of words to their fundamental form through the application of vocabulary and morphological analysis. This methodology aids in the preservation of lexical semantics by simplifying words to their corresponding dictionary form.

During the preprocessing phase, various techniques were employed to train the model on the original cleaned text, stemmed text, and lemmatized text. Stop word removal was not conducted due to the presence of several stop words in our dataset that were considered essential to maintaining the flow and significance of the conversation.

Algorithm:

The problem of the Somaiya Institute's custom dataset is a multi-class classification problem, where the model needs to classify the given prompt into one of 17 classes. To convert the text input into vectors, we used well-known techniques in the NLP field, such as CountVectorizer and TF-IDF.

To address the issue of an imbalanced dataset, we used Synthetic Minority Over-sampling Technique (SMOTE). SMOTE internally uses K-Nearest Neighbors to generate new prompts, thereby balancing the classes in the dataset.

To classify the prompts, we experimented with several machine learning algorithms, including RandomForest, MultinomialNB, SGDClassifier, and KNeighborsClassifier. We used every combination of the three preprocessing techniques (original cleaned text, stemmed text, and lemmatized text), two embedding techniques (CountVectorizer and TF-IDF), and four machine learning algorithms.

We stored the best score for each combination, and the best combination was used for further predictions. This approach helped us to find the most effective combination of techniques and algorithms to achieve the best results.

Interface:

To set up the interface for our chatbot, we used Google's Dialogflow, which provides an easy-to-use interface for developing chatbots. We also created a Flask application that serves as the backend for our chatbot.

A webhook was created to link the Dialogflow interface to our Flask server. We used Ngrok to expose the localhost link externally, making it accessible to users over the internet.

We set the prompt setting to fallback intent, which enabled us to pass the prompt directly from Dialogflow to our Flask server. The prompt was then preprocessed using the techniques described earlier, and the preprocessed prompt was passed through our machine learning model. The model classified the prompt into one of the 17 tags, and the appropriate response was sent back to Dialogflow via the response webhook.

Dialogflow provides an iframe that can be embedded on any website or platform, making it easy to integrate our chatbot with a variety of platforms.

Result:

The results of the experiment showed that the combination of stemming, TF-IDF, and MultinomialNB achieved the highest accuracy of 90.9%. The other algorithms also achieved reasonable accuracies, with the RandomForest algorithm achieving an accuracy of 74%, the SGDClassifier achieving an accuracy of 89.7%, and the KNeighborsClassifier achieving an accuracy of 78.7%.

Conclusion:

This project aimed to develop a chatbot that could assist users with frequently asked questions related to the Somaiya Institute. We generated a custom dataset consisting of 17 tags and used various NLP techniques such as tokenization, stemming, and lemmatization to preprocess the text input. We then used CountVectorizer and TF-IDF to convert the text input into vectors and handled the problem of an imbalanced dataset using SMOTE.

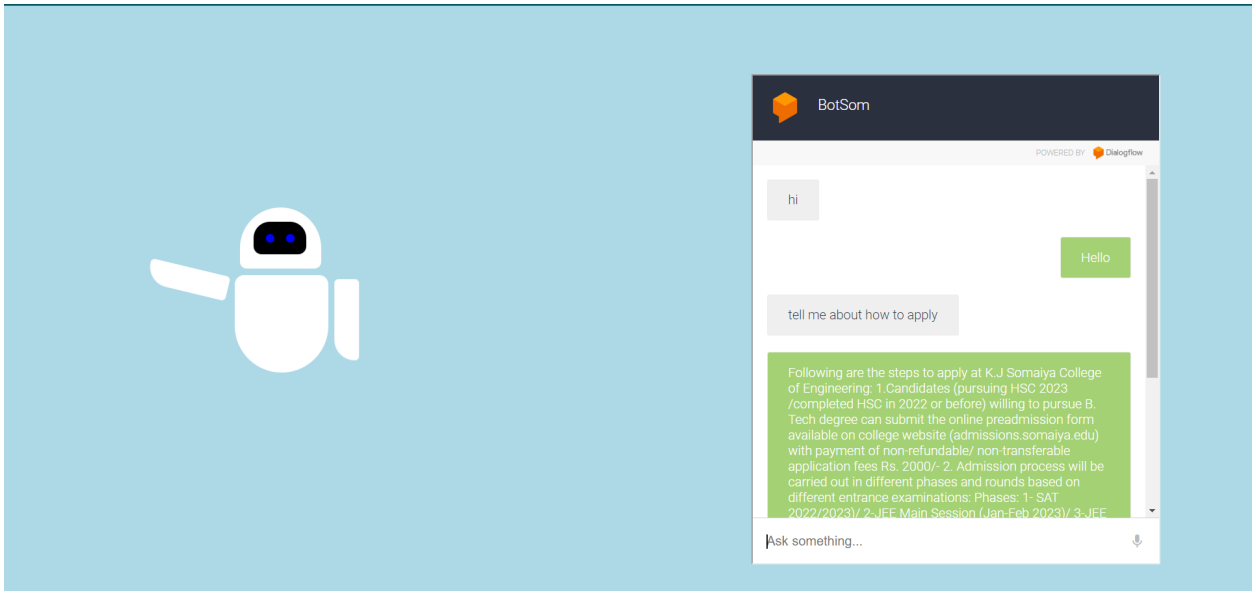
Next, we used four different machine learning algorithms to classify the prompt into one of the 17 tags. We trained and tested the model using various combinations of preprocessing techniques and embedding techniques and found the best combination to be using the stemmed text input along with TF-IDF vectorization and the MultinomialNB algorithm. Finally, we integrated the model with a chatbot interface using Google's Dialogflow and a Flask application as the backend.

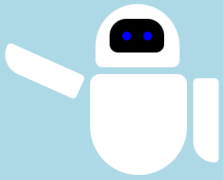
Our approach achieved excellent results in terms of classification accuracy and can help reduce staff workload while providing a better user experience for the students and faculty members of the Somaiya Institute. In conclusion, our project demonstrated the feasibility of using a chatbot to assist users with frequently asked questions and showcased the power of combining various NLP techniques and machine learning algorithms to achieve accurate and efficient results.

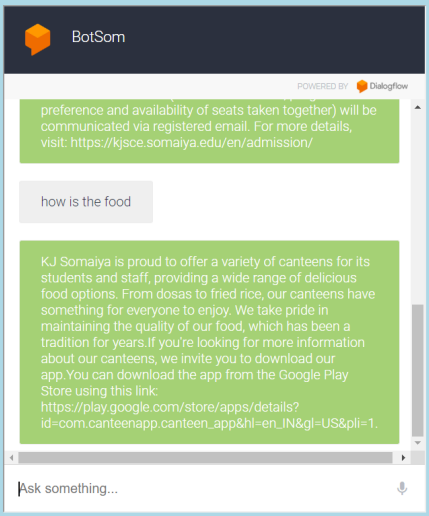
Github Link:

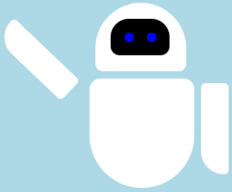
https://github.com/kashish1211/BotSom-Somaiya-ChatBot

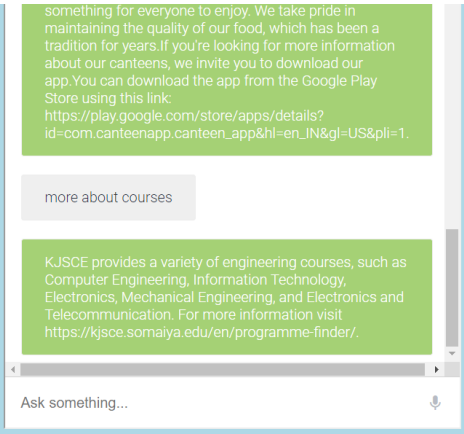
Screenshots:

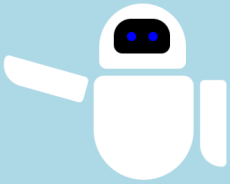


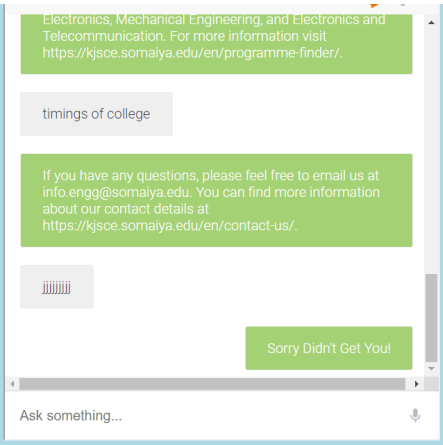



BotSom
POWERED BY Dialogflow
preference and availability of seats taken together) will be communicated via registered email. For more details, visit: https://kjsce.somaiya.edu/en/admission/
how is the food
KJ Somaiya is proud to offer a variety of canteens for its students and staff, providing a wide range of delicious food options. From dosas to fried rice, our canteens have something for everyone to enjoy. We take pride in maintaining the quality of our food, which has been a tradition for years.If you're looking for more information about our canteens, we invite you to download our app.You can download the app from the Google Play Store using this link: https://play.google.com/store/apps/details?id=com.canteenapp.canteen_app&hl=en_IN&gl=US&pli=1.
Ask something...




something for everyone to enjoy. We take pride in maintaining the quality of our food, which has been a tradition for years.If you're looking for more information about our canteens, we invite you to download our app.You can download the app from the Google Play Store using this link: https://play.google.com/store/apps/details?id=com.canteenapp.canteen_app&hl=en_IN&gl=US&pli=1.
more about courses
KJSCE provides a variety of engineering courses, such as Computer Engineering, Information Technology, Electronics, Mechanical Engineering, and Electronics and Telecommunication. For more information visit https://kjsce.somaiya.edu/en/programme-finder/.
Ask something...




Electronics, Mechanical Engineering, and Electronics and Telecommunication. For more information visit https://kjsce.somaiya.edu/en/programme-finder/.
timings of college
If you have any questions, please feel free to email us at info.engg@somaiya.edu. You can find more information about our contact details at https://kjsce.somaiya.edu/en/contact-us/.
jjjjjjjj
Sorry Didn't Get You!
Ask something...

References:

- https://medium.com/@divalicious.priya/creating-a-chatbot-using-the-basic-ml-algorithms-part-1-70f6af52c2e3
- https://chatbotsmagazine.com/chatbot-development-with-nlp-a-beginners-guide-c146115dd657
- https://towardsdatascience.com/using-nlp-to-improve-chatbot-performance-lessons-from-industry-leaders-8f83464880bd
- https://medium.com/analytics-vidhya/building-a-chatbot-with-natural-language-processing-5d5e5df5cf5e
- https://ieeexplore.ieee.org.library.somaiya.edu/document/8576921
- https://www.researchgate.net/publication/354044170_Investigating_the_user_experience_of_customer_service_chatbot_interaction_a_framework_for_qualitative_analysis_of_chatbot_dialogues
- https://www.analyticsvidhya.com/blog/2020/01/how-to-build-your-first-chatbot-using-nlp/