



Educational Institute based Chatbot using Natural Language Processing

P Chandra Sekhar^{1,*}, G Haritha², S Anusha³, M Nithya Sree⁴, I Kavitha⁵

Department of Computer Science and Engineering, Aditya College of Engineering, Madanapalle, Andhra Pradesh, India

* Corresponding Author: P Chandra Sekhar; harithgajala59@gmail.com

Abstract : Chatbot for educational institutions that makes use of natural language processing (NLP) and neural network techniques. The chatbot is a creative way for educational institutions to improve support and communication. Through the utilization of neural networks and natural language processing, the chatbot enables smooth communication between staff, faculty, and students. It provides individualized support for questions about academic resources, scheduling, course details, and administrative processes. Because of its neural network architecture, the chatbot can comprehend and provide responses that are similar to those of a human, which enhances user engagement and experience. The chatbot learns continuously, improving its ability to understand context, tailor its responses, and adjust to changing linguistic patterns. Sentiment analysis and intent recognition are two NLP approaches that enable the chatbot to quickly and accurately evaluate user requests and deliver pertinent information.

Keywords : Machine Learning, NLP, Data Collection, AI, Data Science, Anaconda, Navigator, Django, CSS.

1. Introduction

A computer program known as a chatbot helps people have lifelike conversations with one another. The continuous progress in communication and information technologies has led to an increase in the complexity of artificial intelligence. Artificial intelligence systems are designed to replicate human behaviors.

Examples of these behaviors include responding to users promptly, handling routine tasks, making judgments at predetermined periods, and answering questions in a way that is congruent with human decision-making. There are a lot of electronic companies out there, including those in the e-business, virtual assistance, entertainment, and other sectors. Everything has an internet connection these days. Having everything delivered directly to your home is a highly efficient method to manage and reap the benefits of it all. Users can be led to believe they are conversing with a real person by the sheer number of chatbots available.

2. Literature Survey

Generals communicate using words, gestures, and facial expressions. Thanks to social media's ease of use, they may also express themselves through stories on Instagram and WhatsApp as well as through tweets. We attempted to

examine the fundamentals of user-posted tweets in this piece; the dataset is available here.

First, we determine if the class distributions in the train and test datasets are identical or different. It is a good idea to verify the distribution before moving further with model construction since there are situations when differences in the class distribution between the test and train sets could have an impact on how well the model performs on test or unseen data.

Additionally, understanding the distribution of classes in your data will assist you in selecting an appropriate metric for your issue. For example, if your dataset contains imbalanced classes, accuracy might not be the best metric to employ as accuracy might lead to deceptive conclusions.

For instance, if your model is stupid and labels every sample as positive when you have 90 samples from the positive class and 10 samples from the negative class, you would have an accuracy of 90%, which presents a misleading picture of the model's performance.

The distribution of labels in the train and test datasets is similar, and the data is reasonably balanced, as the figure below illustrates.

3. Technical Description

With the goal of making package management and deployment easier, Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.). The "Conda" package management system is in charge of managing package versions.

More than 12 million people use the Anaconda distribution, which comes with more than 1400 well-liked data science programs that work with Windows, Linux, and MacOS.

It is therefore unnecessary to learn how to install each library separately because the Anaconda distribution includes more than 1,400 packages in addition to the Conda package and virtual environment manager known as Anaconda Navigator.

With the conda install command or the pip install command that comes bundled with Anaconda, each open source package may be installed separately from the Anaconda repository. Many of the functionality found in conda packages are also available in pip packages, and they may usually be used in tandem.

The conda build command can be used to create custom packages, which can then be uploaded to PyPI, Anaconda Cloud, or other repositories for distribution. Python 2.7 and Python 3.7 are included in the Anaconda2 and Anaconda3 default installations, respectively. On the other hand, you can build new environments with any Python version that comes included with conda.

Python is a high-level, general-purpose programming language that is interpreted. Its design concept uses a lot of indentation to enhance code readability. Its object-oriented methodology and linguistic constructions are intended to assist programmers in writing logical, understandable code for both small and large-scale projects.

Python uses garbage collection and dynamic typing. It is compatible with various programming paradigms, such as object-oriented, functional, and structured (especially procedural). Because of its extensive standard library, it is frequently referred to as a "batteries included" language.

Python was developed by Guido van Rossum in the late 1980s as an alternative to the ABC programming language. Python 0.9.0 was initially made available in 1991. With the release of Python 2.0 in 2000, new capabilities including list comprehensions and a reference-counting garbage collection mechanism were included. 2008 saw the release of Python 3.0, a significant update to the language that is

Jack Sparrow Publishers © 2024, IJRDES, All Rights Reserved
www.jacksparrowpublishers.com

not entirely backwards compatible. In 2020, Python 2 was phased out with version 2.7.18.

Python was developed as a replacement for the ABC programming language, which was influenced by SETL and was able to handle exceptions and interface with the Amoeba operating system, in the late 1980s by Guido van Rossum at Centrum Wickenden & Informatica (CWI) in the Netherlands. It started to be implemented in December 1989. As the lead developer, Van Rossum bore full responsibility for the project until July 12, 2018, when he declared his "permanent vacation" from his duties as Python's Benevolent Dictator for Life.

The Python community conferred this title on him in recognition of his enduring dedication as the project's principal decision-maker. A five-person "Steering Council" was chosen in January 2019 by active Python core contributors to oversee the project.

The current members of this council as of 2021 are Pablo Galindo Salgado, Thomas Wouters, Carol Willing, Barry Warsaw, and Brett Cannon.

4. Conclusion and Future Work

This chatbot's primary goal was to create an algorithm that would recognize the user's questions or queries and respond appropriately. to create a database where all pertinent information is kept and cross-referenced with inquiries as they come up.

We have effectively created a chatbot that allows parents or students to ask questions about admissions, eligibility requirements, course specifics, and the inquiry process. The chatbot evaluates the query and provides the appropriate answer.

We can make our college inquiry chatbot more interactive in a variety of languages in the future to better serve consumers from diverse geographical areas. In addition to offering an answer, the future chatbot must to offer a solution to the parent's or student's issue.

In addition to information on admissions, the College Enquiry Chatbot's electronic copy should also include details about academic achievements, placements, and scholarships.

References

- [1]. H. Li, A. Mukherjee, B. Liu, R. Kornfield, and S. Emery, "Detecting campaign promoters on twitter using markov random fields," in Proc. IEEE Int. Conf. Data Mining, 2014, pp. 290–299.



- [2]. R. Mourtada and F. Salem, "Civil movements: The impact of facebook and twitter," Arab Soc. Media Rep., vol. 1, no. 2, pp. 1–30, 2011.
- [3]. J. Ratkiewicz et al., "Truthy: Mapping the spread of astroturf in microblog streams," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 249–252.
- [4]. J. Ratkiewicz, M. Conover, M. R. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in Proc. 5th Int. AAAI Conf. Weblogs Soc. Media, 2011, pp. 297–3040
- [5]. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in Proc. 18th Int. Conf. World Wide Web, 2009, pp. 551–560.
- [6]. C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," Nat. Commun., vol. 9, no. 1, pp. 1–9, 2018.
- [7]. K. Thomas, C. Grier, and V. Paxson, "Adapting social spam infrastructure for political censorship," in Proc. 5th USENIX Conf. Large-Scale Exploits Emergent Threats, 2012, pp. 1–8. [Online]. Available: <https://www.usenix.org/conference/leet12/workshop-program/presentation/thoma>.