

Design and Development a MediBot Disease Prediction System using Apriori algorithm and RNN

Tathagat Yadav, Ankit yadav

Assistant Professor, Department of Computer Science & Engineering, Sapthagiri College of Engineering, Bangalore, India

Article Information

Received : 02 June 2024
Revised : 06 June 2024
Accepted : 18 June 2024
Published : 22 June 2024

Corresponding Author:

Tathagat Yadav

Abstract— In recent years Machine Learning has a major impact in the field of medical science due to its ability to learn and analyze from the examples provided. There are lots of treatments that are available for various common diseases that can be detected by individuals and doctors just on the basis of symptoms. But it is not possible for any human to know about all diseases. So, the problem is that there isn't any place where anyone can have the details of the diseases based on the symptoms. What if there is a place where we can find our health problem just by entering symptoms. The proposed idea is to create a chatbot that can predict disease using Apriori algorithm just by providing symptoms. Also, by using Recurrent Neural Network algorithm this system can be simply used as a tool of communication. Hence it can be beneficial for people as early detection of the disease can help in better treatment and recovery. Moreover, this system will help people to keep track of their health regularly and properly without going anywhere.

Keywords: *Apriori Algorithm, Chatbot, Disease Prediction, ML, NLP, RNN*

Copyright © 2024: Tathagat Yadav, This is an open access distribution, and reproduction in any medium, provided Access article distributed under the Creative Commons Attribution License the original work is properly cited License, which permits unrestricted use.

Citation: Tathagat Yadav, "Design and Development a MediBot Disease Prediction System using Apriori algorithm and RNN", Journal of Science, Computing and Engineering Research, 7(6), June 2024.

I. INTRODUCTION

In today's fast paced life people often don't take proper care of their health. People try to avoid going to doctor for diagnosis either because they don't like it or they don't get time. Now let's imagine you have a friend who is a doctor then your life would be so easy. You can randomly have a conversation with your friend about your health and your friend responds by asking more about the symptoms which you are having and will then diagnose you for a disease which you are suffering from. But not everyone can have a friend who is a doctor and as a result people often end up ignoring their health conditions. Well not anymore. In today's era where technologies like Apple's Siri, Microsoft Cortana is at the forefront of the highly personalized virtual assistants. They can do things which people imagined few decades ago. Similar to these technologies we can build a Chatbot where everyone can have a doctor friend with whom he can talk in general as well as talk about health issues. Chatbots are programs that mimic human conversation using Machine Learning algorithms. It is designed to be the ultimate virtual assistant helping one to complete tasks ranging from answering questions, getting driving directions, to playing one favorite tune etc. Chatbots has become more popular in business groups right now as they reduce customer service cost and handles multiple users at a time. They are being used in almost every domain from virtual assistant like Siri on mobile to customer support in tech industry to e-commerce websites. Chatbots are currently the one of the best trending technologies available. It is certainly one of the most advanced and time saving

technology also. But yet to accomplish many tasks there is a need to make chatbots efficient in medical field as well. To address this problem this project provides a platform where human can interact with a chatbot which are highly trained on datasets using Machine Learning algorithms. Machine Learning algorithms take a more natural approach for computation rather than taking a logical approach. The output is depended on the dataset they are trained on. Even without knowing the proper logic behind an algorithm one can make use of those algorithms and can be benefited from it. It is like learning from the experiences. But in Machine learning the biggest challenge is the choice of algorithm that is to be used for the dataset. It is said that all Machine Learning algorithms are wrong but some are useful. There are number of models available for machine learning. For the proposed chatbot the current best industry algorithms are used.

II. RELATED WORKS

A survey on the techniques that can be used to design Chatbots is made [5]. The different design techniques of nine papers that can be used to design a chatbot are compared. It also covers the significant improvements in Chatbots in the last decade. In addition, it discusses about the similarity and differences in the techniques that are used in the selected study and then with the Loebner Prize winning chatbots techniques. The proposed work presents the methods of teaching chatbots to process natural language text. It discusses about processing natural language using Recurrent Neural Network (RNN). The sequence to sequence long short-term

memory cell neural network (LSTM) is used to train the model. In addition, it also talks about the challenges of implementing a Recurrent Neural Network based chatbot [1]. Disease diagnosis system using several machine learning algorithms is proposed. A detailed comparison of four Machine Learning algorithms to predict disease based on symptoms provided, is also presented in this paper. The purpose was to find the best Machine Learning algorithm to diagnose diseases early to help doctors and patients, as early prediction of disease can improve treatment efficiency [12]. The work proposed here discusses the need of a chatbot on e-commerce site to enhance user experience. The chatbot gives information about the products on an e-commerce sites which is useful for a customer to buy exactly what they want. The chatbot is designed using Artificial Intelligence Markup Language (AIML). It highlights the multi-lingual properties of the chatbot. The Graphical User Interface (GUI) is developed in Bangla Language. The proposed work also discusses the limitations of AIML Knowledge Based System to process natural language [11]. A problem solving chatbot using Deep Neural Network is proposed [4]. The chatbot is trained to remember different Data Structures such as array, queue, stack and tree. It can take inputs in text or voice format and respond form a knowledge base. The output is based on the learning provided during the training of the chatbot. The main purpose is to show that chatbot can also be used for problem solving.

III. METHODOLOGY

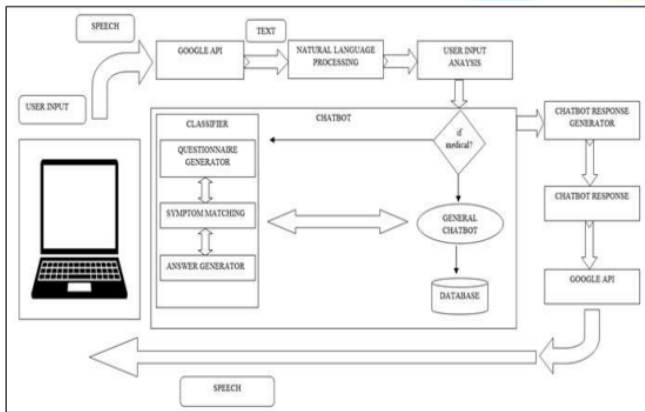


Fig. 3(a): System Architecture

In the proposed system we have tried to enhance the ability of a chatbot from being used just as a tool of communication to predict disease based on symptoms provided as input. To achieve this, the system combines various machine learning algorithms. The system design is as follows: User Interact with Chabot The user starts a conversation with the chatbot by providing input through speech. The input is recorded through a microphone. It is just like starting a normal conversation with any human being. The conversation with the chatbot is completely voice based. Decision Making As soon as the chatbot starts the conversation with the user, the

user is asked to choose from the options provided. Two options are provided: General Conversation and Medical Assistant. Based on the response from the user the chatbot decides if it has to continue a general conversation or to help the user for self-diagnosis. Response from chatbot The response is generated based on the decision making in the previous stage. If the user needs help related to health the chatbot enters into a questionnaire. It starts asking about different symptoms and then finally predicts a disease. If the user does not need any help then the chatbot continues with the normal conversation. All the response from the chatbot is voice based.

IV. RESULTS AND DISCUSSIONS MODULE DESCRIPTION

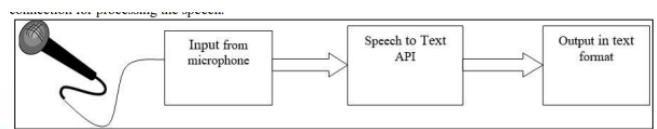


Fig. 4(a): Speech-to-Text Processing

Speech-To-Text In this module the user voice input is being converted into text. This is to be done because text input is required for the next stage. We will be using a Google Speech Recognition API for Speech-to-text conversion. We can speak in the microphone of our laptops and the Google API will translate it into text. This Google API is very powerful and it can also translate several other languages into text. But for this proposed system only English is required. However, this API does not work offline and it requires a working internet connection for processing the speech.

Natural Language Processing The computer system cannot understand the natural language of the humans. Hopefully we have a branch of computer science, called Natural Language Processing, which deals with the natural language which human speak. The goal of NLP is to take the unstructured output of the API and produce a structured representation of the text that contains spoken language understanding. The output of the Google API, which will be a text input, is given as input to our chatbot system. The chatbot uses Natural Language Processing to process the text input. After processing of the text input, the chatbot responds with a series of questions in order to understand the situation of the user better. A number of methods are available to extract meaning from spoken words. We have used Bag of Words. In Bag of Words sentence structure order and syntax is ignored only the count of number of occurrences of each word is taken into consideration. This is used to form a vector space model in which unnecessary words are removed. It is used to extract meaning information from the spoken words. Response generation The response generation is the most important phase of the chatbot system. This chatbot will respond based on the requirements of the user. The interaction with the chatbot is classified in mainly two types: General Response and

Medical Assistant. General Response The general response will be created based on the training of the chatbot. There are number of ways for training the chatbot. In this system we have used the Sequence-to-Sequence (seq2seq) model to train the system. The model has been trained on the Cornell Movie Corpus dialogue dataset. Medical Assistant If the user wants any medical diagnosis of the disease based on the symptoms provided, then the chatbot uses the Apriori Algorithm to predict the disease. The model is trained on dataset available from the New York Presbyterian Hospital. This algorithm is best suited for frequently occurring item sets in the dataset. Text-to-Speech After the response generator analysis and text input it generates an output according to the user input. The output can either be general conversational speech or it can be a questionnaire of symptoms to predict disease. The output from the response generator has to be converted from text to speech in both the cases. For this purpose, the Google text-to-speech API is used to generate a voice output from the chatbot. The API is capable of responding in a number of languages but for this project only English is used.

V. ALGORITHMS USED

Two algorithms are used in this proposed system. They are: Sequence-to-Sequence Model The seq2seq model consists of two RNN: an encoder and a decoder. The encoder takes a sentence as input and processes one word one at a time. The decoder generates words one by one in each time step of the decoder's iteration. After one complete iteration, the output is generated.

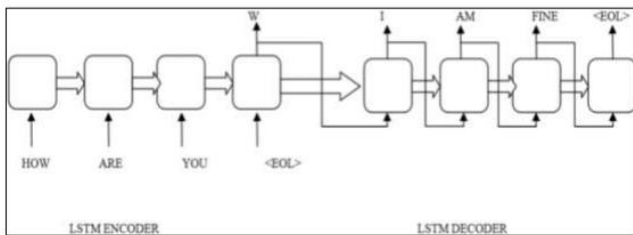


Fig. 5(A): Seq2Seq Model

Apriori The Apriori algorithm is used for finding frequent item sets in a dataset for Boolean association rules. The apriori principle can reduce the number of items sets we need to examine. The algorithm uses bottom up approach where frequent subsets are extended one at a time, known as candidate generation and group of candidates are tested against the data. It states that if an item set is infrequent, then all its supersets must also be infrequent. This means that if {pale eyes} was found to be infrequent, we can expect {pale eyes, cold} to be equally or even more infrequent, so in consolidation the list of popular item sets, we need not consider {pale eyes, cold}, nor any other item set configuration that contains pale eyes.

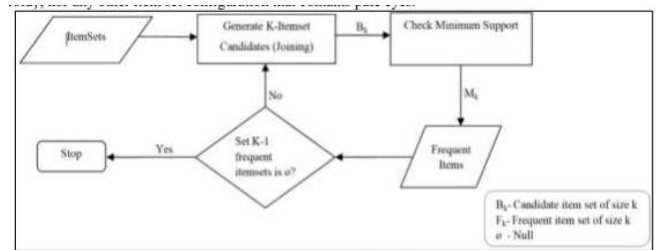


Fig. 5(b): Apriori Algorithm

VI. CONCLUSION

A chatbot design was implemented which can be used as a tool for general conversation as well as for self-diagnosis. The system can be very useful in the field of medical science for early and faster detection of disease. Also, the system helps an individual to keep track of their health more properly. However, the biggest challenge is the lack of correct and accurate medical dataset. Although there are a number of dataset available for developing a conversational chatbot, there is only one dataset available for disease prediction which is limited. There is one more big challenge that the seq2seq model requires a lot of time for training even though the hardware is capable of handling it. For future enhancements we can use an offline Application Programming Interface (API) for Speech-to-text conversion and Text-to-Speech conversation. Also, the symptom-diseases mapping can be done more accurately if a more reliable dataset is made available. Along with this, with the increasing use of smart wearable more accurate body measurements like heart rate, Body Mass Index (BMI) etc. can be given as input to the system and the disease prediction algorithm can be improved for more accurate prediction of diseases.

REFERENCES

- [1]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1955009, 13 pages, 2022. <https://doi.org/10.1155/2022/1955009>.
- [2]. D. Sathyanarayanan, T. S. Reddy, A. Sathish, P. Geetha, J. R. Arunkumar and S. P. K. Deepak, "American Sign Language Recognition System for Numerical and Alphabets," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369455.
- [3]. J. R. Arunkumar, Taguele berihun Mengist, 2020" Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier" *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.

- [4]. Ashwini, S., Arunkumar, J.R., Prabu, R.T. et al. Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network. *Soft Comput* (2023). <https://doi.org/10.1007/s00500-023-09480-3>
- [5]. J.R.Arunkumar, Dr.E.Muthukumar," A Novel Method to Improve AODV Protocol for WSN" in *Journal of Engineering Sciences*" ISSN NO: 0377-9254 Volume 3, Issue 1, Jul 2012.
- [6]. R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [7]. Arunkumar, J. R. "Study Analysis of Cloud Security Challenges and Issues in Cloud Computing Technologies." *Journal of Science, Computing and Engineering Research* 6.8 (2023): 06-10.
- [8]. J. R. Arunkumar, R. Raman, S. Sivakumar and R. Pavithra, "Wearable Devices for Patient Monitoring System using IoT," 2023 8th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India, 2023, pp. 381-385, doi: 10.1109/ICES57224.2023.10192741.
- [9]. S. Sugumar, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [10]. R. S. Vignesh, V. Chinnammal, Gururaj.D, A. K. Kumar, K. V. Karthikeyan and J. R. Arunkumar, "Secured Data Access and Control Abilities Management over Cloud Environment using Novel Cryptographic Principles," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199616.
- [11]. Syamala, M., Anusuya, R., Sonkar, S.K. et al. Big data analytics for dynamic network slicing in 5G and beyond with dynamic user preferences. *Opt Quant Electron* 56, 61 (2024). <https://doi.org/10.1007/s11082-023-05663-2>
- [12]. Krishna Veni, S. R., and R. Anusuya. "Design and Study Analysis Automated Recognition system of Fake Currency Notes." *Journal of Science, Computing and Engineering Research* 6.6 (2023): 16-20.
- [13]. V. RamKumar, S. Shanthi, K. S. Kumar, S. Kanageswari, S. Mahalakshmi and R. Anusuya, "Internet of Things Assisted Remote Health and Safety Monitoring Scheme Using Intelligent Sensors," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199766.
- [14]. R. S. Vignesh, R. Sankar, A. Balaji, K. S. Kumar, V. Sharmila Bhargavi and R. Anusuya, "IoT Assisted Drunk and Drive People Identification to Avoid Accidents and Ensure Road Safety Measures," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10200809.
- [15]. I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [16]. G. Karthikeyan, D. T. G, R. Anusuya, K. K. G, J. T and R. T. Prabu, "Real-Time Sidewalk Crack Identification and Classification based on Convolutional Neural Network using Thermal Images," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1266-1274, doi: 10.1109/ICACRS55517.2022.10029202.
- [17]. R. Meena, T. Kavitha, A. K. S, D. M. Mathew, R. Anusuya and G. Karthik, "Extracting Behavioral Characteristics of College Students Using Data Mining on Big Data," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10084276.
- [18]. S. Bharathi, A. Balaji, D. Irene, J. C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.
- [19]. I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [20]. Revathi, S., et al. "Developing an Infant Monitoring System using IoT (INMOS)." *International Scientific Journal of Contemporary Research in Engineering Science and Management* 6.1 (2021): 111-115.