

Doctor Talkbot

A Study of Different Conversational Systems

Deepak S¹, Dr M.N. Nachappa²

¹Post Graduate Student, School of Computer Science and Information Technology, JAIN (Deemed to be University), Bangalore deepaks24june@gmail.com

²Professor and Head, School of Computer Science and Information Technology, JAIN (Deemed to be University), Bangalore mn.nachappa@jainuniversity.ac.in

ABSTRACT

Conversation agents, often known as chatbots, are software programmes that can converse with humans as if they were two humans. Artificial intelligence is used by these bots to understand human input and reply appropriately. Many people in today's environment rely heavily on their smartphones. They could employ chatbots for personal support and information extraction from the internet. The medical/health, agricultural, and educational realms are all vital to consider. Chatbots may now execute duties such as reduced agent transfers, faster issue resolution, increased self-service, and more with little downtime and cost savings. Other chatbots carry out prediction jobs (particularly in the medical field), which are now achievable because to advances in AI and data mining techniques. The purpose of this article is to perform a survey to compare existing chatbots in terms of platforms, algorithms, tools and software, and other factors.

Keywords—Medical Chatbot, Heart Disease Prediction, Healthcare, Data Mining, Support Vector Machine (SVM)

I. INTRODUCTION

Chatbots are computer programmes that can converse in a natural manner. This technology is gaining in favour, particularly in the commercial world, and it has the potential to cut personnel. Some employ natural language processing techniques, while others just scan the input for keywords and return a response with the most relevant keywords/patterns from the database. E-commerce (through chat), customer assistance, education, entertainment, finance, food, games, health (Medical), HR, marketing, news, and travel are just a few of the categories where chatbots can be found. Chatbots can assist users in locating their requirements as data grows. ELIZA was the first chatbot ever constructed. Alexa and Siri are two examples of chatbots that have been developed to date.

A medical chatbot is a programme that can diagnose a problem, assess its severity, and provide treatments based on the symptoms presented to the system. It can instantly connect a patient with a doctor in an emergency. Chatbots that can take user input (questions), recognise it, and respond with relevant output can be created using AI and data mining.

People avoid going to the hospital for minor issues. In the future, little issues will turn into serious diseases. The early detection of serious diseases is difficult. Going to the hospital might take a long time and cost a lot of money. The healthcare business has a wealth of data, but due to a lack of resources, little of it is used.

E-commerce chatbots ensure that your customers buy your products directly from the conversation itself in order to cope with their difficulties. Customers can now ask natural questions regarding the things they've been looking at. It assists clients in locating the appropriate products by applying the necessary filters/sorts and receiving prompt responses, and then purchasing them directly from the chat tool.

The goal of this study is to evaluate different chatbots that take in necessary data from users as input, process it using data mining, and produce necessary output, while also understanding the necessity for conversational systems in various domains and viewing the challenges associated with such domains.

II. LITERATURE SURVEY

The purpose of the literature review or background was to examine and learn from past studies and surveys. Some studies were categorised based on the tools/software they used, the methods they employed, and the datasets they utilised (if any) in conjunction with the platform. There are also some survey papers describing the comparison of many existent chatbots.

Using Cleveland and Statlog Heart Datasets in MATLAB R2010, algorithms such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) were used to predict heart diseases, with the resulting accuracy, precision, and sensitivity shown in table 1 [1]. Natural Language Understanding, Word Embedding Models, Framework for Emotional Recognition, Spatial-Temporal Context Analysis, Multi-modal Approach for defining an Emotional Expression Model to Categorize, and Collecting Training Data for Emotion Recognition are all used to define an Emotional Expression Model to Categorize, Continuous Emotional Monitoring using data, in text, image, video, audio format [2].

SWITCHes Android app uses mHealth Intervention Tools to construct a custom medical chatbot algorithm for Android under the LINE channel for Weight Control and Health Promotion using the energy balance equation [3]. Under the medical domain, accuracy of 98.3 percent can be obtained using AI, Deep Learning, and Machine Learning algorithms (with the Tensorflow library) to deliver personalised healthcare and fitness-related assistance [4]. MATLAB R2012a was used to compare Genetic Algorithm, Decision Trees, and Naive Bayes Algorithms, and it was utilised to construct a prototype that can discover and extract unknown knowledge connected to heart disease [5].

MySQL, AIML, swftools, Red5, HTTP, RTMP, and a Single-Supply Biasing Method were all employed in the MediAssistEdge system. The main system was made up of two sub-systems: the DocBot system and the MediConnect system, both of which simplify diagnostics and improve patient-doctor connectivity [6]. A chatbot created for cancer sufferers or anyone interested in learning more about cancer receives the user's message over social media platforms such as Facebook, WhatsApp, and others. The chatbot scrapes data from cancer forums and stores it in a local database using the Beautiful Soup python library. It also uses the NLTK python library for pre-processing data. The Responder processes the retrieved

information and generates a human-like output [7]. VDMS, a web-based chatbot, was created using program-o, an open-source PHP interpreter tailored for AIML. VDMS's local knowledge base on diabetes is built using AIML pattern tags. Pattern matching algorithm is also used. A test was done in which a group of ten people addressed questions to VDMS, and the chatbot was able to answer 65 percent of the questions to the satisfaction of the users [8].

MedBot makes use of datasets from DoctorMe Apps, Natural Language Understanding (NLU), Dialogflow API, and Pattern matching algorithm. With the proper APIs, the chatbot may be simply integrated into online chat platforms like Facebook, Hangouts, or Line. The competency of the proposed system is determined by the test output (results) [9]. The WEKA tool and KEEL were used to analyse the Decision Tree Algorithm with Cleveland Clinic Foundation (Cleveland Data) for Heart Disease Prediction, with an accuracy of 87.41722 percent accurately identified instances. ECR, cholesterol, chest pain, fasting blood sugar, maximum heart rate reached, and other important measures were used. The missing values in the datasets were filled in using the Allpossible-MV method [10]. In the dialogue manager, AI and machine learning algorithms were contrasted utilising the MAS (Multi-Agent System) support tool kit to achieve anticipated behaviour and RL (Reinforcement Learning) to make learning easier. RL [11] was used to train the bot using datasets from the knowledge base.

After distinguishing hyperplane that minimise the error for unseen patterns, SVM, k-NN, Naive Bayes, and Porter Stemming Algorithms, as well as AIML, Google API, Java & Pascal, MySQL, NLP, and particular Heart Diseases datasets, were utilised to develop a Heart illness prediction system [12]. The test for heart disease detection was decreased by applying the feature-subset selection method, which selects useful data from a dataset while removing irrelevant data. The genetic algorithm employs a genetic search to reduce the number of attributes in the data set for a more straightforward search. Sellapanetal's dataset of record 909 with 13 attributes was used [13].

MedChatbot is an open-source AIML-based Chatterbean that answers queries using the widely known UMLS knowledge store. The AIML-based chatbot is built in such a way that natural language inquiries are converted into SQL queries. These queries are then executed against/with the knowledge base, yielding natural dialogue responses [14]. Farmbot is an agriculture-focused chatbot that employs natural language processing (NLP) techniques to comprehend a user's natural language enquiry. The bot is trained using a training dataset, a neural network is created based on the dataset, and the errors are optimised using the gradient descent algorithm. To give voice-based responses to the user, the chatbot uses a "speech synthesis web API." The chatbot also employs a prediction algorithm known as "ARIMA" to forecast future agricultural commodity costs, which when tested yielded a mean absolute percentage inaccuracy of 0.1814 percent for the data [15]. A chatbot created with the Eclipse software on a Windows operating system, using the Java programming language. Java applets were utilised to create a chat dialogue window. The database for this chatbot was created by using a two-dimensional string array. This chatbot had a minimal design and only answered questions if they were discovered in the database [16].

TABLE I. TABLE OF COMPARISONS OF SOME RESEARCH PAPERS

Ref. No.	Methodology	Strengths	Weaknesses	Accuracy	Domain	Voice/Text
[1]	Data pre-processing and Training & Testing the Models in SVM & ANN.	The Testing Procedure for System Model was regarding Specification, Precision, Accuracy, and Sensitivity.	Multiple Thresholds need to be added for Progressing.	(precision, accuracy, sensitivity) SVM – 85.6%, 84.7%, 84.1% ANN – 83.3%, 81.8%, 68.7%	Medical (Heart)	Text
[2]	Defines an Emotional Expression Model in order to Categorize, Collect Training Data for Emotion Recognition, Emotion Recognition, and Inference, Continuous Emotional Monitoring.	Focuses on human emotion recognition and its monitoring, and also conversation understanding on chat assistant.	It needs to expand this conversational service to game addiction cases of teenage, and thereby investigate clinical results with user satisfaction.	67.52% on avg. for all seven emotional states.	Mental/ Psychiatric Counselling	Both
[3]	Uses energy balance equation, which is: $E_s = E_i - E_e$.	SWITCHes app was designed to take in an individual's dietary intake and energy expenditure data on a regular basis into the webserver.	-	-	Weight Control	Text
[4]	Built using Convolutional Neural Network using Tensorflow.	Great way to deliver personalized healthcare and fitness information.	The user has to jump into various services for healthcare and fitness	98.39%	Medical	Text
[5]	It's a prototype which can extract and determine unknown knowledge related with heart disease from a past heart disease records.	Cost-effective computer-based treatment. Support Systems can be developed to make better decisions.	Bad clinical decisions can lead the death of a patient which can be expensive for any hospital	Naive Bayes – 90% and Decision Tree- 96.5%	Medical (Heart)	Both
[6]	Single-supply biasing method	DocBot system can provide easy access, 24/7 availability and anonymity to the users.	When a disease affects a crowd of people, regional doctors can get overburdened and cannot consult all patients at the same time.	-	Medical	Both
[7]	The chatbot makes use of Beautiful Soup python library to scrap data from cancer forums and store it in a local database also uses NLTK available in python for pre-processing of data.	Cancer bot helps Cancer patients to share their problems regarding cancer when they are in a helpless situation. It can learn from previous conversational experience.	Accuracy of the chatbots generated responses depends on the accurate data scrapped from cancer forums.	-	Medical	Text
[8]	Uses open-source PHP interpreter designed for AIML called program-o to build its core. It uses MediaWiki API if the user's query match is not found in the local database to explore from external Wikipedia knowledge source.	Program-o is open-source technology and has a flexible structure which can accommodate code modifications.	AIML doesn't have a logic engine and has a poor pattern matching functionality with the presence of fixed knowledge in the chatbots knowledge-base.	65%	Medical	Text
[9]	Uses Dialogflow Platform	Dialogflow supports Natural Language Understanding which can develop chatbots without writing code.	Chatbot may respond with the incorrect answers because of small training phrases.	Not Mentioned	Medical	Both
[10]	Uses Decision tree Algorithm for prediction of Heart Diseases.	Divides data set into 2 parts for performance evaluation.	There can be possibilities of classification errors.	Correctly Classified Instances 87.41722%	Medical	-

[11]	Made a conceptual model connecting the system to the information source. Used MAS tools to achieve anticipated behavior in the dialog manager, and RL to ensure facilitate learning.	Has access to reliable and accurate maternal health information. Helps in reducing waiting time in hospital avoiding queues.	Don't have any consolidated health record systems and may rely/feed on another system.	-	Medical	Text
[12]	The SVM classifier discovers the finest distinguishing hyperplane which minimizes the error rate for unusual patterns. NLP & Word Order Similarity among Sentences.	SVM classifiers are faster to train w.r.t the best feature space size. Hence performance is increased for small or medium sizes.	Customer Satisfaction are the Major Concerns for developing this System.	SVM - 94%, Naive Bayes - 80%, KNN - 88%	Medical (Heart)	Both
[13]	Feature subset selection, Naïve Bayes, Decision tree, clustering	Prediction of heart diseases are more accurate with lesser number of attributes.	Clustering classification performed poorly as compare to other methods	Naïve bayes =96.5%, Decision tree=99.2%, Clustering =88.3%	Medical	-
[14]	MedChatbot uses open-source and AIML based Chatterbean and also uses common UMLS as the knowledge source to answer the queries.	Acts as an interactive tutoring system for Medical Students.	Wasn't able to answer queries for which UMLS knowledgebase didn't contain any information.	-	Medical	Text
[15]	Chabot uses Natural Language Processing technique to understand the user queries in natural language. It also makes use of prediction algorithm "ARIMA"	Provides an interface where farmers or other users can interact effectively and get the desired responses with limited steps.	The dataset contains hundreds of agricultural queries. If the user's queries are not found, then the chatbot will struggle to provide the desired response.	0.181% Mean Absolute Percentage Error of the prediction	Agriculture	Both
[16]	Creating the dialog box, Creating a database to store various type of questions the user may ask, and responses that a Chatbot can answer.	Simple Implementation of chatbot. Easy to implement.	Works only on windows operating system.	-	E-Commerce	Text

A survey discusses artificial intelligence, chatbot history, and chatbot applications. With an increase in scripts produced for it, new ways outperform the pattern matching algorithm. From the very first chatbot to the most recent ones, a comparison survey was conducted on the existing approaches, design, and implementation of many chatbot systems. Implementation of a domain-specific chatbot, Implementation of a smart responding Optical Character Recognition based chatbot, and Implementation of an Inquisitive chatbot were the three methodologies used. Table 2 shows the results of the comparison.

TABLE II. EXISTING CHATBOT COMPARISON SURVEY

Chatbot & Year	About	Features
ALICE 1995	Is a program that uses some heuristical pattern matching rules for engaging in a conversation with a human with input given.	Uses AIML, is an open-source, text-based inquisitive chatbot.
SIRI 2011	Computer program that works as an intelligent personal assistant, part of Apple Inc. operating systems.	Uses Java, JavaScript, Objective C, NLP, TTS, STT. It is a Self-learning, voice & text-based Inquisitive chatbot.
ALEXA 2015	Its an Intelligent personal assistant which is developed by Amazon Lab126, capable of voice based interaction, music playback, etc.	Uses NLP, TTS, STT, Python, Java, and Node.js. It is a Self-learning voice-based Inquisitive chatbot.
MITSUKU 2012	Mitsuku claims to be a female Chatbot from Leeds who is 18-years old.	Uses AIML, is Self-learning, text-based and Inquisitive.
ELIZA 1964	A simulation of a Rogerian psychotherapist, rephrasing her response with a few grammar rules.	Uses MAD-Slip, is an open-source & text-based chatbot.
Jabberwacky 1982	Chatbot having an aim that states: "simulate natural human conversation in an interesting and entertaining manner".	Uses Contextual pattern matching, learned by association, storing replies to inputs in a database. It is a self-learning text-based chatbot.

PARRY 1972	PARRY could simulate the behavior of a person having paranoid schizophrenia.	It is a complex system of assumptions, attributions, and “emotional responses” triggered by weights assigned to verbal text inputs.
WATSON 2006	It is a question answering system which had won the Jeopardy contest in 2011.	It uses IBM’s DeepQA and Apache UIMA. It is a self-learning & text-based Inquisitive chatbot.

III. PROPOSED METHODOLOGY

The following tools and technologies can be utilised for better chatbot design, analysis, and accuracy, as studied in the literature survey in the previous section, in order to create/build a chatbot or conversational system for any domain (general):

A. Technology and Tools

1) Datasets: On platforms like Kaggle, Github, and others, there are several free datasets available for various subjects. They might be available on other websites as well. For example, there are three databases accessible for diagnosing cardiac disease. The data was acquired from three locations: Cleveland Clinic Foundation, Hungarian Institute of Cardiology, University Hospital, Switzerland, and Hungarian Institute of Cardiology, University Hospital, Switzerland.

2) ML/DM Algorithms: After evaluating the data, algorithms such as Support vector machine (SVM) can be utilised for classification and regression analysis. One of the classes is assigned to each training dataset. The SVM training algorithm is a binary linear non-probabilistic classifier that creates a model that allocates new instances to one of two classes. This algorithm was chosen because it has a higher accuracy rate than other algorithms.

3) Dialogflow (API.ai): This is Google's free Platform that implements interactive voice-based and text-based conversational interfaces (using NLP) driven by AI, giving users better ways to connect with their service or product. Dialogflow uses Google's machine learning skills and technologies like Google Cloud Speech-to-Text . Basic working of Dialogflow is shown in fig. below.

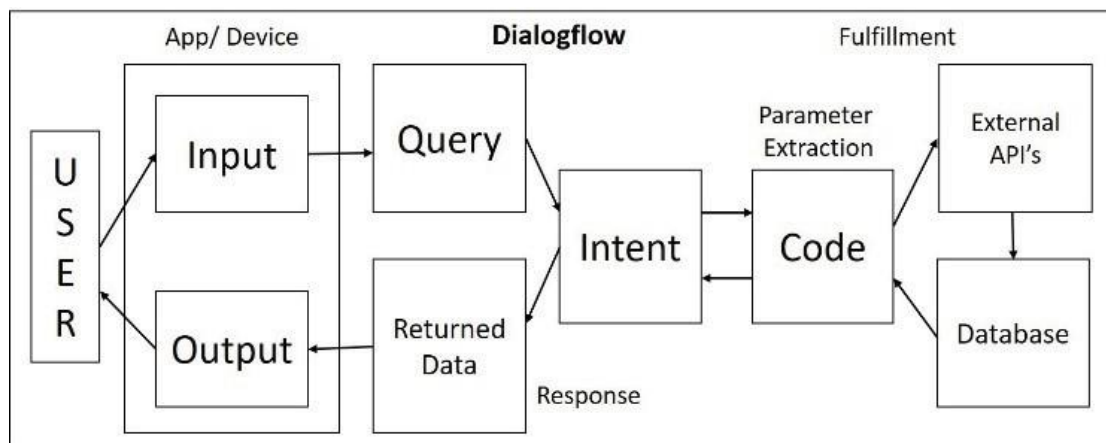


Figure 1: Shows a working diagram of a basic Dialogflow.

4) MATLAB: This is a numerical computing IDE that allows you to manipulate matrices, plot data, implement algorithms, create user interfaces, and interface with programmes written in other languages.

5) WEKA 3: Weka is an open-source Data Mining tool/software that employs a collection of machine learning techniques that may be applied directly to data. It also includes tools for data pre-processing, clustering, classifications, and visualisation.

6) AIML: Artificial Intelligence Mark-up Language is a type of XML that can be used to build conversational agents. Some chatbots plan to respond to user messages using AIML, which can create accurate and relevant responses. It's simple to programme, easy to comprehend, and easy to maintain.

B. Discussion

Dialogflow is an excellent framework for developing chatbots. It can recognise the provided query and respond appropriately using machine learning and intent matching approaches. AIML, an easy-to-understand programming language, can also be used to create chatbots. The Questions and Answers are written in XML format and saved. The Support Vector Machine algorithm clusters linear and non-linear data from a dataset, resulting in more accurate results, particularly in the medical arena. Datasets that are freely available to the public can be used in such an algorithm. Data analysis, graph displaying, and algorithm implementation can all be done with tools like WEKA and MATLAB. WEKA is a data mining tool that may be used to pre-process data, cluster it, and categorise it. Chatbots may thus be constructed for a variety of fields, including medical, e-commerce, and education, using these tools and technology.

IV. CONCLUSION

Most implemented chatbots use AI and Data Mining algorithms such as SVM, Decision tree, Genetic algorithm, Naive Bayes, Pattern Matching, NLP and more using tools such as WEKA or Matlab for analysis or execution. Some algorithms were tested and compared based on accuracy after giving input datasets. Some chatbots used AIML for the QA structure. This paper specifies Dialogflow as the Conversation Interface. Under Medical domain, the data mining model which gives the best accuracy for predicting these disease from heart disease dataset was the Support Vector Machine algorithm.

V. REFERENCES

- [1] Mrs. S. Radhimeenakshi; "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network"; 978-9-3805-4421-2/16/\$31.00 ©2016 IEEE, 2016 International

- Conference on Computing for Sustainable Global Development (INDIACom).
- [2] Kyo-Joong Oh, DongKun Lee, ByungSoo Ko, Ho-Jin Choi; "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation"; 2375-0324/17 \$31.00 ©2017 IEEE, 2017 IEEE 18th International Conference on Mobile Data Management.
 - [3] Chin-Yuan Huang, Ming-Chin Yang, Chin-Yu Huang, Yu-Jui Chen, Meng-Lin Wu, Kai-Wen Chen; "A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion"; 978-1-5386-6786-6/18/\$31.00 ©2018 IEEE, Proceedings of the 2018 IEEE IEEM.
 - [4] Siddhant Rai, Akshayanand Raut, Akash Savaliya, Dr. Radha Shankarmani; "Darwin: Convolutional Neural Network based Intelligent Health Assistant"; 978-1-5386-0965-1/18/\$31.00 ©2018 IEEE, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).
 - [5] Ankita Dewan, Prof. Meghna Sharma; "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification"; 978-9-3805-4416-8/15/\$31.00 ©2015 IEEE, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom).
 - [6] Amiya Kumar Tripathy, Rebeck Carvalho, Ajit Puthenpuhussery, Nikita Chhabhaiya, Bijoy Anthony; "MediAssistEdge – Simplifying diagnosis procedure & Improving patient doctor connectivity"; 978-1-4799-8187-8/15/\$31.00 ©2015 IEEE, 2015 International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04–06, 2015, Mumbai, India.
 - [7] Belfin R V, Ashly Ann Mathew, Blessy Babu, Shobana A.J., Megha Manilal; "A Graph Based Chatbot for Cancer Patients"; 978-1-5386-9533-3/19/\$31.00 ©2019 IEEE, 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS).
 - [8] Shafquat Hussain, Prof. Athula Ginige; "Extending a conventional chatbot knowledge base to external knowledge source and introducing user based sessions for diabetes education"; 978-1-5386-5395-1/18/\$31.00 ©2018 IEEE, 2018 32nd International Conference on Advanced Information Networking and Applications Workshops.
 - [9] Nudtaporn Rosruen and Taweesak Samanchuen; "Chatbot Utilization for Medical Consultant System"; 978-1-5386-7573-1/18/\$31.00
 - [10] Purushottam, Richa Sharma, Prof.(Dr.) Kanak Saxena; "Efficient HeartDisease Prediction System using Decision Tree"; ISBN:978-1-4799-8890-7/15/\$31.00 ©2015 IEEE, International Conference on Computing, Communication and Automation (ICCCA2015).
 - [11] K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375. doi: 10.1109/MDM.2017.64
 - [12] Du Preez, S.J. & Lall, Manoj & Sinha, S. (2009). An intelligent web-based voice chat bot. 386 - 391.10.1109/EURCON.2009.5167660
 - [13] Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling", Intelligent Systems Modelling and Simulation (ISMS) 2016 7th International Conference on, pp. 72-77, 2016.
 - [14] Dahiya, Menal. (2017). A Tool of Conversation: Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING. 5. 158-161.2017.
 - [15] C.P. Shabariram, V. Srinath, C.S. Indhuja, Vidhya (2017). Ratatta: Chatbot Application Using Expert System, International Journal of Advanced Research in Computer Science and Software Engineering,2017
 - [16] Mrs Rashmi Dharwadkar1, Dr.Mrs. Neeta A. Deshpande, A Medical ChatBot, International Journal of Computer Trends and Technology (IJCTT) – Volume 60 Issue 1- June 2018
 - [17] Farheen Naaz, Farheen Siddiqui, modified n-gram based model for identifying and filtering near-duplicate documents detection, International Journal of Advanced Computational Engineering and Networking, ISSN: 2320- 2106, Volume-5, Issue-10, Oct.-2017
 - [18] N-gram Accuracy Analysis in the Method of Chatbot Response, International Journal of Engineering & Technology. (2018)