Building a Medical Chatbot using Support Vector Machine Learning Algorithm

To cite this article: B. Tamizharasi *et al* 2020 *J. Phys.: Conf. Ser.* **1716** 012059

View the article online for updates and enhancements.

Recent citations

Building a Medical Chatbot using Support Vector Machine Learning Algorithm

Tamizharasi B., Jenila Livingston L.M.* and S. Rajkumar

School of Computer Science and Engineering, VIT, Chennai, India - 600127

E-mail:*jenila.lm@vit.ac.in

Abstract. Chatbot is used extensively to check the state of health at any time. It is the same as going to a doctor and having the medication prescribed. This paper discusses about medical chatbot using the machine learning algorithm which predicts the accuracy of the disease. There are many machine learning algorithms that can be used to predict the disease. Support Vector Machine learning technique is primarily used to achieve precise prediction and boost the efficiency of the model. The system uses Natural Language Processing to achieve the style of chatting. Using this approach people can reduce spending time in hospitals and receive low cost or cost-free services.
Keywords: Medical chatbot, Machine Learning Algorithm, Support Vector Machine algorithm, Natural Language Processing.

1. Introduction

In this world, all that we want is perpetual joy and happiness. We experience happiness and joy through our sense organs in the body. We feel happy only when we maintain proper physical health. If health deteriorates our happiness declines which will lead to pain and misery. Nowadays folks are less attentive to their health that results in high risk once it's not renowned.

A Medical chatbot plays a major role in health and provides correct diagnosing. The user will discuss concerning their symptoms the chatbot predicts the sickness and recommends the treatment. Building chatbots has emerged as an honest approach for medical organizations. Medical chatbots will perform one-on-one interactions with patients accessing the information management network of the organization. Counting on the accuracy and details their script, they are in the addition able to verify each patients individual complaints, requests and adopt approach to provide the assistance they need. This methodology is continual in any form of times and conjointly the online chatbots to learn and improve their interaction.

1.1. Machine Learning

Ability of a system to learn itself, and improve based on the learning without comprehensive programming is known as Machine learning which is broadly classified into 4 types. They are Supervised, Unsupervised, Semi-supervised and Reinforcement methods (Figure 1).
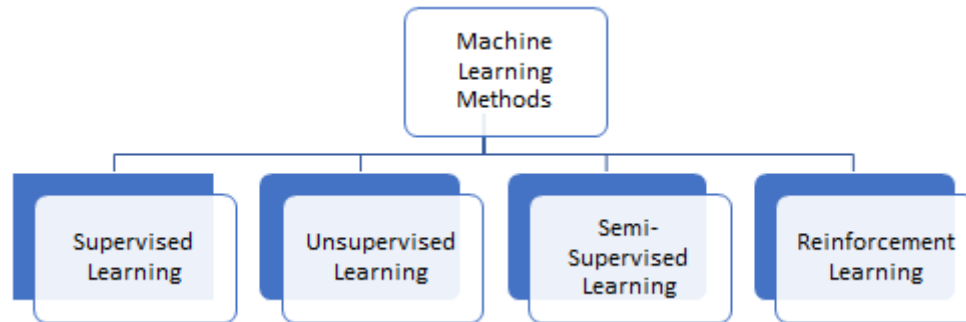
Figure 1. Classification of ML Algorithms

Machine learning uses programmed algorithms to predict output values within an appropriate range of input data received and analysed. These algorithms help in solving various problems in the IT sector effectively.

1.1.1. Supervised Learning Technique

The aim is to collect data and learn from it using various machine learning based supervised techniques. It is a mathematical model consisting of set of inputs (labelled inputs) and desired outputs used for predictive modelling. Some commonly used algorithms are Nearest Neighbor, Decision tree, Support vector machines, Naïve Bayes and linear regression.

1.1.2. Unsupervised Learning Technique

In this method, the data are unlabelled; the system trains itself and produces the output. This helps in picking the important data that is required for the analyses. When we require any information regarding the relationship of a data, we use unsupervised learning method. Some commonly used algorithms are association rule and K-means clustering.

1.1.3. Semi-Supervised Learning Technique

It is a combination of supervised and unsupervised learning methods where we will use both labelled and unlabelled data. The basic process involved is that first the programmer would use an unsupervised learning algorithm to cluster related data, and then use the current labelled data to mark the remaining unlabelled data. It produces a desired result having important parameters required for analyses. Semi-supervised learning's practical applications are speech analysis, Internet content classification and sequence classification of proteins:

1.1.4. Reinforcement Learning Technique

This method is trained based on the trial and error method for a particular decision. Reinforcement learning is all about sequential decision making. Simply put we can say that the output depends on the current input status and the next input depends on the output of the previous input. That is, it gains experiences from the previous trainings and gives accurate knowledge based on the response received.

2. Related Work

The model suggested by Divya Madhu et. al. [1] using AI that facilitates individuals to establish the correct treatment for patient's disease. There are several treatments on the market for a selected disease and nobody will probably grasp the correct treatment that is best fitted to their disease. Here the main intension is to use AI, which predicts the disease accurately with the symptoms. AI system predicts the correct illness and shows treatments.

Divya et. al. [2] proposed self-diagnosis of medical chatbot using artificial intelligence techniques. Number of illnesses and treatment and diagnosis appeared in the database. There will be several illnesses and therapies so that AI system predicts the disease correctly and reliably. AI

checks the age and symptoms, then forecasts the data and provides the correct advice and diagnosis.

Amiya et. al. [3], developed the Doc-Bot to predict the disease and prescribe medicine. A small application was developed here to say patient's symptoms. At the most, user should create an account with the user name and password and a unique id is given. Moreover, the queries will be asked and required blank should be filled where age is the main criteria to fix the disease of the algorithm used here. Once the disease is found by them it can be clarified by others review and multiple questions can be asked to them to clarify the efficacy of the drug and availability.

Rashmi and Neeta [4] have developed a medical chatbot that uses the google API which converts the voice-text to text-voice and is used to speak dramatically like a virtual doctor. Here NLP is used to fulfil this technology which changes the interface. Hence a patient-doctor virtual platform is created to establish cost free services for knowing the medicine and diagnosis. The proposed system is developed based on their work.

Ranoliya et. al. [5] proposed a model using API interface which is most commonly used to create a chatbot with voice text conversion and to create a virtual platform of a hospital. It clearly uses the FAQs for diagnosing the diseases where a number of queries are asked and then treatment is decided. After that medicine is given which may be again checked by feedback and questions for supplying accurate medicines.

Sagar, Jenila and Agnel [6] developed a prediction system using an ML algorithm to predict tourist attraction spots and their associated information by analysing social media data which in turn helps the tourism industry to understand what kinds of experiences visitors can have and how their desires can be met. Algorithms such as KNN, Kernel Density Estimation, Random Forest, and XG Boost have been used for this purpose. The findings showed that XG Boost produces performance with higher accuracy than other three algorithms.

The proposed approach by Lucas et. al. [7] used leaf spectral data to determine the amount of macro and micro nutrients measured in the laboratory. The accuracy is obtained using the KNN, ANN, and RF algorithms, of which RF performed the best. Imran and Shikha [8] implemented Artificial Intelligent Mark-up Language (AIML) based chatterbot to train the model, and used Microsoft voice synthesizer to identify a user's spoken word. Natural language processing is used for Microsoft speech comprehension and recognition.

Pryss et. al. [9], proposed an architecture that would serve as a reference point for various psychological and medical scenarios. This research puts forward aspects that could form the basis for better collaboration between medical as well as technical experts for proper chatbot solutions.

Xuewen, Xiaoping, John [10] recommended a SVM training algorithm for human and bots classification in text-based communication chatbot using annual Loebner competition data. At each conversations, approximate normalized entropy and inter message delays is imposed. The simulation results showed that the SVM is an effective method for the classification of chatbot data.

3. Proposed Model

This system helps users submit complaints and queries regarding their health. The main concern regarding the development of this system is user satisfaction. The actual motive of the chatbot is to encourage people by offering appropriate advices on safe and healthy living. The overall system architecture is given in Figure 2.
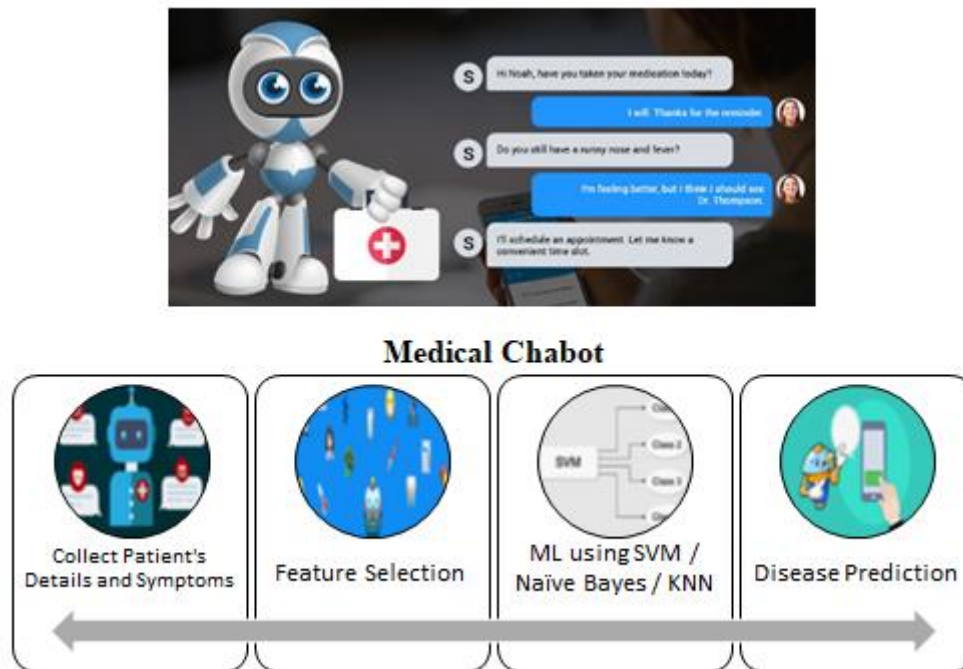
Figure 2. System Architecture

3.1. Working Mechanism

The code is written in JAVA and is an android application. Characteristics of this chatbot are, it collects basic personal details, symptoms and other medical related details. Data will be reviewed with dataset using the Support Vector Machine (SVM) learning algorithm once the symptoms are given. The dataset will be in AIML format, where checks can be made faster. The chatbot will then locate the exact illness and suggest remedies.

3.2. Support Vector Machine

SVM is a supervised learning algorithm and is applicable for linearly separable data. For non-linear data kernel functions are used. SVM classifies the two classes using "hyper plane". The hyper plane has the highest margin for separating given data into classes. SVM is a regulated AI calculation which can be utilized for both arrangement and relapse difficulties. In this method, we plot every data as a point in n-D space with the number of features with the estimation of each element being the estimation of a specific range. Classification is being performed at this point to find the hyper-plane which separates the two classes. In SVM, it is anything but difficult to have a direct hyper-plane between these two classes. SVM has a method called the Kernel. Its capacities that take up low-dimensional information space and transform it into a higher-dimensional space. It is helpful in classifying non-directional issue.

3.3. Chatbot Conversation

The similarity of word order between two sentences is significant, since the meaning of a sentence may be dramatically altered by different word order. The mechanism of the chatbot is, Once the chatbot asks for name and age it will check whether age is below 18 or not. If the age is below 18 the chatbot will stop the conversation by saying them to inform their parents else the conversion continues. The sample chatbot interaction between human and machine is depicted with Figure 3.

Figure 3. Sample Chatbot Conversation

4. Experimental Results

This medical chatbot is trained with different datasets with 70% training dataset and 30% testing dataset. Sample dataset is specified in Figure 4.

Figure 4. Sample Dataset Used

For simple testing KNN and Naïve Bayes algorithms are also used. Mostly SVM gives accurate results and it worked better with large number of data and working faster too. The comparative analysis is given in Table 1 and depicted with the column chart (Figure 5). SVM produces 92.33%, KNN with 87.66% and Naïve Bayes with 81%.

Table 1. Comparative Analysis of Classification Algorithms

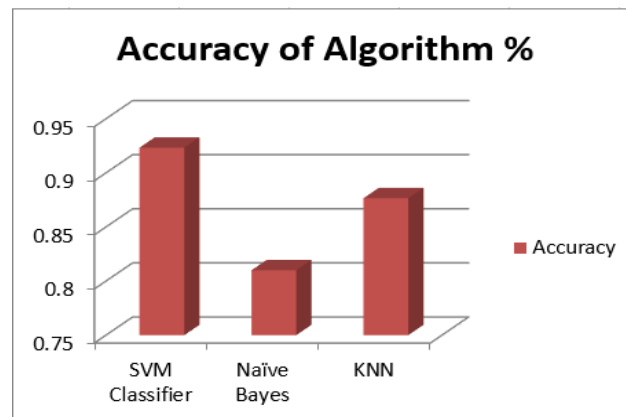| Algorithm Comparision			
Sr.No	Algorithms	No of Disease conditions	Accuracy
1	SVM Classifier	200	0.92333
2	Naïve Bayes	200	0.81
3	KNN	200	0.8766

Figure 5. Classification Accuracy of Machine Learning Algorithms

Hence it is clearly proved that SVM classifier is used to predict the disease accurately compared to other ML algorithms.

5. Conclusion

The proposed system benefits medical institutes and hospitals to assist patients through voice or text queries to freely raise medical related inquiries. The system obtains output from the diagnosis of medical API and speaks out with remedies to the disease. SVM classification accuracy is higher comparable with KNN and Naïve Bayes algorithms. SVM produces 92.33% accuracy and is thus used to accurately predict the disease compared to other machine learning algorithms and also saves time and space.

6. Future Enhancement

By taking the advantages of the SVM algorithm medical chatbots can be extended and used deeply with other medical systems where predictions can be done. It can be further extended to schedule doctor visits and remind patients of a next appointment or routine check-up. Also it can be extended to collect patients' feedback and this will help medical organizations to improve their processes.

References

[1]   Divya Madhu, Neeraj Jain C J, Elmy Sebastain, Shinoy Shaji and Anandhu Ajayakumar 2017 A Novel Approach for Medical Assistance Using Trained Chatbot *International Conference on Inventive Communication and Computational Technologies (ICICCT)* pp 243-246.

[2]   Divya S, Indumathi V, Ishwarya S, Priyasankari M and Kalpana Devi S 2018 A Self-Diagnosis Medical Chatbot Using Artificial Intelligence *Journal of Web Development and Web Designing* vol **3** (1) pp 1-7.

[3]   Amiya Kumar Tripathy, Rebeck Carvalho, Keshav Pawaskar and Suraj Yadav 2015 Mobile based healthcare management using artificial intelligence *International Conference on Technologies for Sustainable Development (ICTSD)* pp 4-6.

[4]   Rashmi Dharwadkar and Neeta A Deshpande 2018 A Medical Chatbots. *International Journal of Computer Trends and Technology* vol **60** (1) pp 41-45.

[5]   Ranoliya B R, Raghuvanshi N and Singh S 2017 Chatbot for university related FAQs *International Conference on Advances in Computing Communications and Informatics (ICACCI)* pp 1525-1530.

[6]   Sagar Gupta, Jenila Livingston L M and Agnel Livingston L G X 2019 Prediction of Top Tourist Attraction Spots using Learning Algorithms *International Journal of Recent Technology and Engineering* ISSN: 2277-3878 vol **8** (3) pp 1063-1067.

[7]   Lucas et al. 2020 A Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements *Remote Sensing* MDPI vol **12** (6) 906 doi:10.3390/rs12060906 pp 1-21.

[8]   Imran Ahmed and Shikha Singh 2015 AIML Based Voice Enabled Artificial Intelligent Chatterbot *International Journal of u-and e-Service, Science and Technology* vol **8** (2) pp 375-384.

[9]   Pryss R et al. 2019 Using Chatbots to Support Medical and Psychological Treatment Procedures: Challenges, Opportunities, Technologies, Reference Architecture. Baumeister H, Montag C (eds) Digital Phenotyping and Mobile Sensing *Studies in Neuroscience, Psychology and Behavioral Economics* Springer, Cham, ISBN 978-3-030-31619-8 pp 249-260.

[10]   Xuewen Mu, Xiaoping Shen and John Kirby 2017 Support Vector Machine Classifier Based on Approximate Entropy Metric for Chatbot Text-based Communication *International Journal of Artificial Intelligence* ISSN: 0974-0635 vol **15** (2) pp 1-16.