Implementing Heuristic Rule Model based Chatbot using Natural Language Processing - Result Paper

[1]Shashank Dilip Sable, [2]Prof. Suvarna Lehekar, [3]Dr. V B Kamble

[1]M.Tech. Student, [2]Professor, [3]Head of Department
PES College of Engineering, Aurangabad

***Abstract*: A chatbot is a computer software that communicates with users of messaging systems by using artificial intelligence (AI). Every time a user enters information into the chatbot, it remembers it and answers. As a result, the chatbot, which had no prior information, might learn from the replies received. A chatbot system may be put up online to assist website visitors. With the aid of this application, we may easily access files without having to navigate through several modules. artificial intelligence-based methods, such natural language processing (NLP). Once set up, chatbots are accessible 24/7 and eliminate the need for human interaction while doing business online. These are able to handle several requests concurrently and automatically update and learn. In the midst of the epidemic, telemedicine has the potential to be useful by allowing patients to receive supportive treatment without having to physically visit a hospital utilising a dialogue artificial intelligence-based software. As a result, telehealth will swiftly and significantly replace in-person therapy with online patient consultation. It developed a global Conversational Bot based on Natural Language Processing (NLP) to provide free fundamental healthcare education, information, and counselling to chronic patients. The study suggests a ground-breaking computer application that acts as a patient's virtual personal doctor. The programme has been meticulously designed and instructed to speak to patients as though they were genuine individuals.**

***Keywords*: Telehealth, chatbot, natural language processing, medbot, natural language understanding, conversational technology, digital health, voice user interface, conversational user interface, conversational agent, human-computer interaction.**

Introduction

One of India's major challenges is providing its growing population with high-quality, cheap healthcare. According to the WHO's World Health Report [1], India ranks 112 out of 190 countries in terms of its healthcare system. Due to the difficulties in obtaining healthcare facilities, particularly in rural India, patients are more inclined to postpone treatment or select facilities that may be nearby but are not cost-effective or well-suited to their medical needs. In the hunt for more efficient ways to provide patients with timely medical treatment, access, and high-quality therapy, the role of telemedicine, which connects patients with healthcare practitioners and healthcare information, comes into play. Due to the current "COVID-19" outbreak, social isolation will last for a very long period in India, especially for individuals who have chronic diseases. The general population will find it more challenging to get healthcare services as a result. The National Health Mission [2] reports that numerous acute illnesses that were detected in India during the lockdown have diminished. Less hospitalisations suggest a shortage of healthcare rather than a rise in disease, according to the statistics, not a decline in disease.

People can benefit from telemedicine in this troubling situation. Conversational artificial intelligence allows medical personnel to diagnose and treat patients without the need for an in-person visit, promoting social distance and reducing the risk of Covid-19 transmission. Chatbots that are driven by artificial intelligence (AI) are displaying the function of a virtual assistant that can manage a communication through speech or text, playing a significant part in the current increasing age of digitization. It utilises voice commands to ask questions, do tasks, and make suggestions based on what the user requests. They can adapt to the individual language preferences, search criteria, and usage patterns of the user with repeated use.

A conversation bot with a voice and/or chat interface can play a vital role in ensuring that primary healthcare is affordable, accessible, and perhaps sustainable in the new digital economy by reducing the existing barriers. Thanks to AI, virtual assistants are now present in every corner of the world. The quick response time and customised user experience greatly assist the deployment of conversational AI for providing Tele-health. Voice assistants use a natural language interface for vocal communication. Voice technologies need to be tailored for the healthcare sector in order to be successful [3]. The two primary voice assistant users in the healthcare sector are patients and physicians.

Doctors utilise these software to collect and record patient data. From the standpoint of the patient, virtual assistants with AI skills that can offer 24x7 care to a variety of patients are a less expensive choice.

Such powerful virtual assistant technologies would be especially helpful to those with chronic diseases, people with impairments, and people who live in distant or rural areas. These systems provide a variety of advantages, such as reduced physician workload, improved patient data security, and on-demand access to medical data, making healthcare accessible and affordable for everyone with a simple user interface [4]. In this study, the use of chatbots in telemedicine is addressed.

Our solution offers a multilingual voice application built on NLP for chronic patients and expectant mothers in need of prenatal care. This is accomplished by turning the user's speech into text, which is then processed and understood by AI. After then, the output is spoken again and sent to the user. Information on the most prevalent diseases in rural India is included in our programmes, with a focus on women's healthcare. Our programme imitates the services of a doctor by providing healthcare advice, home remedies, symptoms, and dietary recommendations based on location.

Related Work

As we set out to develop a chat-based & useful mobile experience for museums, it's important to keep in mind that the Studio team does not consider itself an expert in chatbots, artificial intelligence, or conversational interfaces. We don't consider ourselves to be chatbot gurus, but we do consider ourselves to be leaders in experience design and creative technology. Yet. At the end of our year-long endeavour, we want to be far closer to the expert side of the chatbot spectrum than the novice side. In light of this, our primary goal at the present is to quickly learn as much as we can from the experts who already work in the field of bots.

The team has been doing a literature assessment of recent thinking and writing on the subject over the past several weeks since benchmarking of existing expertise is a crucial component of formative research. These materials have been collected by us into a best-practice library, which is similar to a chatbot book club. It makes sense to publicly share these tools and acknowledge their importance to the team as we move on with the project because we've committed to having a transparent approach. Even though it is thorough, this literature study is far from finished. This library will grow as the project goes forward. Please contribute any resources you may be aware of that we may have overlooked in the comments section or on Twitter.

Certain text-based human-computer interaction systems have been developed like ELIZA [5] that imitates a psychotherapist, and then PARRY [6] which suggests the thinking of a paranoid patient. Raij et al., [7] conducted two separate experiments where they compared virtual human interactions and with a real human in a medical consultation scenario. Their result shows similarity in virtual and real interactions context.

A few text-based human-computer interaction systems, such as ELIZA [5], which mimics a psychiatrist, and PARRY [6], which proposes the thoughts of a paranoid patient, have been created.

In two different trials, Raij et al. [7] examined interactions between a virtual person and a real human in the context of a medical consultation. Their findings indicate a correlation between virtual and actual encounters.

The work of A. Fadhils et al. [8] demonstrates how intelligent conversational systems may be utilised to communicate with elderly people to gather information and do ongoing health condition monitoring, particularly after hospital release. A medical recommendation system particularly created to engage with users and take on the role of a doctor is presented by Amato et al. [9].

Pharmabot is a paediatric generic medicine consultant chatbot that Comendador et al. [10] offer as a tool for prescribing and providing relevant knowledge about generic medications for children.

[1] The author of this paper discusses how to use NLP to analyse data from social media. Automatic summarization is a three-step process that breaks down long paragraphs into manageable pieces. Using named entity recognition, it is possible to recognise objects, names of people, and locations. The most crucial method for categorising verbs, nouns, and adjectives is speech tagging. Through wordsense disambiguation, the meaning of an entire phrase is verified. All of this basically makes up a web mining system, which includes data collection, preprocessing, indexing, and mining. As a result, various methods were integrated and NLP was effectively used.

[2] This study takes into account the increase in the amount of papers available on the web, and how each item has to be categorised into one or more groups in order for subsequent searches to be simple. To lessen the document's complexity, several pre-processing techniques are mentioned, which also refers to dimensionality reduction. Tokenization of the text being processed marks the beginning of the NLP phase.

The feature selection stage, where the documents are processed so that the learning algorithm can be applied to them, is what causes the documents to be handled in the numeric format. The k-NN and k-means algorithms, which represent supervised and unsupervised classification, respectively, are the learning algorithms that are discussed. [3] This essay focuses on the fundamentals of developing a QA system expert system. In contrast to search engines, this system demonstrates using natural language processing how to construct the ideal response to a query.

An expert system's knowledge base is the most crucial component since it is where the solution is developed and where many crucial factors, such as document or information retrieval, are taken into consideration.NLP is used in this situation since the knowledge base is also where the query or the string entered is tokenized and further analysed.

[4] The methods used to create chat-bots are the main topic of this essay. Additionally, it makes comparisons between various chatbots and the kinds of algorithms that are employed to ensure their correct operation.Parsing, pattern matching, AIML, chat script, relational databases, Markov chains, and linguistic trickery are crucial skills needed to construct a chat-bot. With the use of NLP, the article explores how to build a knowledge base and interact with it. Since speech analysis is the process that makes the chatbot interactive, it plays a significant part in the chatbot. There are several phases involved, including 1) voice to text conversion and 2) word splitting and speech tagging. 3) Using algorithms and strategies that chunk sentences selecting a phrase and generating a response.

Proposed System

The process starts from the input query. This input query will be entered by the student. It may be a single word or a phrase. If this query is in the form of a sentence then the query will be first passed through query modulation process. The NLP taking place in the modulation part follows a particular pipeline A pipeline in NLP is a chain of independent modules, each one taking as an input the output of the module before it. Raw Text -> Tokenization -> Lemmatization -> POS-tagging -> Dependency parsing -> Role labelling Tokenization, Lemmatization and POS tagging will play an important role in the bot's query modulation since there many words in a sentence which are not in its natural form and have to be classified into adjectives, nouns, conjunction and other speech of sentence. So to bring a word into its natural form , the process of stemming and lemmatization come into play . For e.g.: Running->Run Since both mean the same thing and this is how the query is modulated. query modulation also has the responsibility of removing the words that will not help the bot in the further process of document retrieval. E.g.: What are the newton's Laws of motion? The modulator removes such words from the query so that the search within the documents gives better results with optimum methods.

Query Extraction The extraction will be based on N-gram division algorithm. Here n means the no. of words to be considered as a single entity for relating its metadata. They are set of co-occurring words within a defined window. For example, for the sentence "The cow jumps over the moon". If N=2 (known as bigrams), then the ngrams would be: • the cow • cow jumps • jumps over • over the • the moon

If X=Number of words in a given sentence K, the number of ngrams for sentence K would be:

After the division of the query the metadata related to each of the gram is checked from the knowledge base. A knowledge base here is the data warehouse for all the possible questions with their solutions. And it consists of a learner which constantly updates it if any new query or a new solution for an old query is found.Training process involves loading example dialog into the chat bot‟s database. This either creates or builds upon the graph data structure that represents the sets of known statements and responses. When a chat bot trainer is provided with a data set, it creates the necessary entries in the chat bot‟s knowledge graph so that the statement inputs and responses are correctly represented.

NLP - interpreting natural language This field of of study is in short concerned with task of how natu ral language can be processed in such a way that it‟s semantics can be understood or interpreted by a computer program, to then act based on these interpretations. This is an essential part of any ch atbot since it is important to try to understand what a user wants to say in order to produce a suiting answer. Although, this is far from a trivial problem, on the contrary it is very complex since natural language often is very abstract. As a branch to NLP there is a field of study called NaturalLangu age Understanding (NLU). While NLP is concerned with processing natural language to be i nterpreted, NLU focuses on the actual interpretation. Bolinda G. Chowdhur y[5] mentionthree main problems within NLU: The first concerns the human thought process, the second the se mantics of a given input and the third knowledge outside the program, or common knowledge. Document/ Information Retrieval: As we have seen that the documents have to be retrieved and this comes under information retrieval part .The dataset preferred will be Wikipedia dictionary since our chat-bot focuses on the education part Wikipedia can be considered as a general platform. Since there are two possible approaches for the working of the chat-bot which are retrieval based and generative based models.Since the initial working of chat-bot could be Retrieval based but the generative models which have an open domain have been working of lately so and since education being a vast field there are many answers to various questions. There are various steps which need to be followed so that bot can retrieve a perfect document and then a perfect answer for a particular query. Processes: 1) Response retrieval 2) Response ranking 3) Response triggering.

Response retrieval Once the query is modulated ,the utterance of the query is compared with set of documents within the knowledge base which is considered to be set of documents . Now the sentences selected will be in a triplet format and they are :(Sprev,S,Snext) S:This represents the optimized sentence, for which the documents are to be retrieved . Sprev :Represents the previous statement to 'S'. Snext: Represents the next statement to 'S'. Well the context for the given statement has to be taken into consideration since that helps us optimize the search within the knowledge base. In the latter stages it will be clear that the context helps the LSTM network make better prediction of the word. 2. Response Ranking: The ranking measure for an answer can be done through the famous Google"s PageRank algorithm. In short PageRank is a "vote", by all the other pages on the Web, about how important a page is. A link to a page counts as a vote of support. If there"s no link there"s no support. Similarly here in this project the vote will be counted for the best possible answer to a given query. The PR of each page depends on the PR of the pages pointing to it. But we won"t know what PR those pages have until the pages pointing to them have their PR calculated and so on… From original google paper : „PageRank or PR(A) can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalized link matrix of the web' Calculate a page"s PR without knowing the final value of the PR of the other pages. Basically, each time we run the calculation we"re getting a closer estimate of the final value. So all we need to do is remember the each value we calculate and repeat the calculations lots of times until the numbers stop changing much. After ranking is done the knowledge base finds out the cluster of answers for a particular query and search for the best rank answer from the cluster and passes out as the output. 3. Response triggering: Now that the documents have been ranked a classification algorithm mostly for the unlabeled dataset so unsupervised learning comes into play. Now by unlabeled dataset we mean that the documents that have been ranked against the initial query have to classified into different category and on the basis of the number of topics which have to be covered for studying . Now this makes the process for answering to a particular query even easier 3.1 Text clustering Since there are so many words present within a document and every document havemany important words so this is where textual clustering is important.

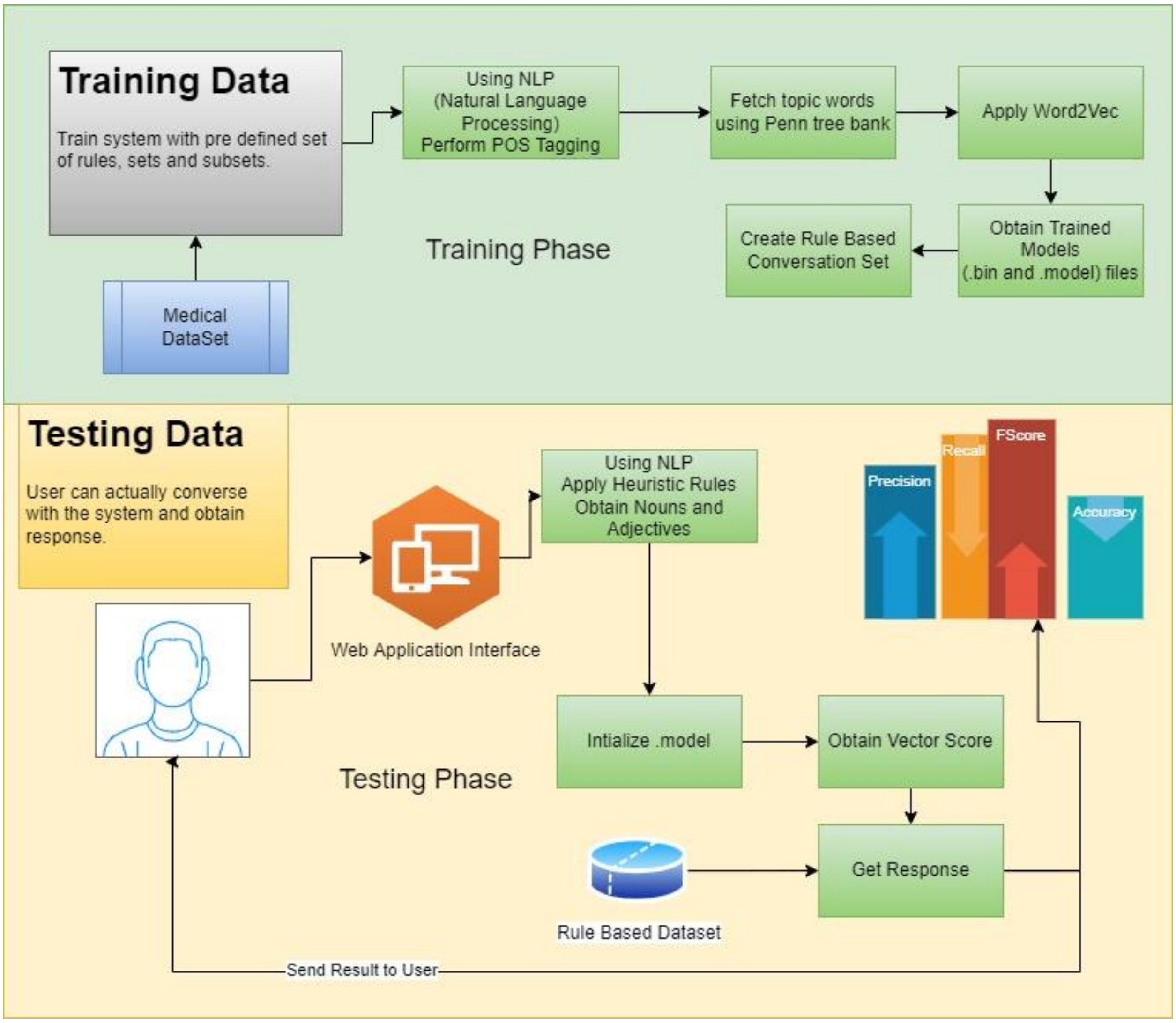

Figure 1.0 Proposed System

1. Word2vec It has two parts which are continuous bag of words and skip gram. Since the continuous bag of words is not our concern so the focus shall be maintained on Skip-gram algorithm. As already discussed for text clustering we need to convert the subjective part into numeric form hence while the bot is coming up with an answer to a relevant query , it will need to vectorize the words from the relevant document and hence to create a proper answer a large corpus would be required hence skip gram method . 1.1 Skip Gram Data sparsity is a large problem in natural language processing and hence skip gram has been used to increase the number of training sentences and hence helps in increasing the size of training corpus. Skip-grams reported for a certain skip distance k allow a total of k or less skips to construct the n-gram. As such, "4-skip-n-gram" results include 4 skips, 3 skips, 2 skips, 1 skip, and 0 skips For e.g.:"Insurgents killed in ongoing fighting." 2-skip-bi-grams = {insurgents killed, insurgents in, insurgents ongoing, killed in, killed ongoing, killed fighting, in ongoing, in fighting, ongoing fighting} 2. Sequence to sequence model: We will be following the sequence to sequence model. Well a sequence to sequence model consist of a pair of RNN(recurrent neural network), the first one works as an encoder which processes the documents and the second one is a decoder which is going to generate an output.

The weights are shared between the encoder and the decoder while they use different set of parameters. The sequence to sequence model takes the input as the vector and hence the skip gram model comes into play. Now the input to the LSTM network shall be from the finalized document, the one that has been ranked the highest and the sentences are selected from it with the contextual reference.

Experimental Setup

Natural Language Processing (NLP) :

NLP techniques are based on machine learning and especially statistical learning which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules . analysis has been handled as a Natural Language

Processing denoted NLP, at many levels of granularity. Starting from being a document level classification task, it has been handled at the sentence level and more recently at the phrase level .NLP is a field in computer science which involves making computers derive meaning from neutral language and input as a way of interacting with the real world.

Feature Extractors

Unigram: Building the unigram model took special care because the Twitter language model is very different from other domains from past research. The unigram feature extractor addressed the following issues:

a. Data contain very casual language. For example, you can search "hungry" with a random number of u's in the middle of the word on http://search.twitter.com to understand this. Here is an example sampling: huuuungry: 17 results in the last day huuuuuuungry: 4 results in the last day huuuuuuuuuungry: 1 result in the last day besides showing that people are hungry, this emphasizes the casual nature of Twitter and the disregard for correct spelling. b. Usage of links. Users very often include links in their data . An equivalence class was created for all URLs. That is, a URL like "http://tinyurl.com/cvvg9a" was converted to the symbol "URL."

Negate as a features

Using the Stanford Classifier and the base SVM classifiers we observed that identifying NEGclass seemed to be tougher than the POS class, merely by looking at the precision, recall and F1 measures for these classes. This is why we decided to add NEGATE as a specific feature which is added when "not" or „n‟t" are observed in the dataset. However we only observed a increase in overall accuracy in the order of 2% in the Stanford Classifier and when used in conjunction with some of the other features, it brought the overall accuracy down and so we removed it. Overlapping features could get the NB accuracy down, so we were not very concerned about the drop with NB. However it didn't provide any drastic change with OpenNLP either.

Part of Speech (POS) features

We felt like POS tags would be a useful feature since how you made use of a particular word. For example, „over‟ as a verb has a positive connotation whereas „over‟ as the noun, would refer to the cricket over which by itself doesn‟t carry any positive or neutrals connotation. On the Stanford Classifier it did bring our accuracy up by almost 6%. The training required a few hours however and we observed that it only got the accuracy down in case of NB Handling the Negatives Class

Word2Vec with Two Classes We extended the Word2Vec Classifier to handle 2 classes: positive, negatives. Collecting a large amount of negatives data is very challenging. For the training data, we simply considered any disease without an emoticon to be part of the negatives class. This is obviously a very flawed assumption, but we wanted to see what the test results would be for the test data, we manually classified 33 data as negatives. The results were terrible. The classifier only obtained 40% accuracy. This is probably due to the noisy training data for the negatives class. 6. Subjective vs. Objective Classifier Another way to handle the negatives class is to have a two phased approach.

Rule Implementation

```
(defrule reply-action
(golden-disease-info (id ?diseaseID)
(text ?diseaseText))
(twitter-user
(language "en")
(screen-name ?screenName))
=>
(assert
(recruitment-action
(action "mention")
(diseaseID ?diseaseID))))
```

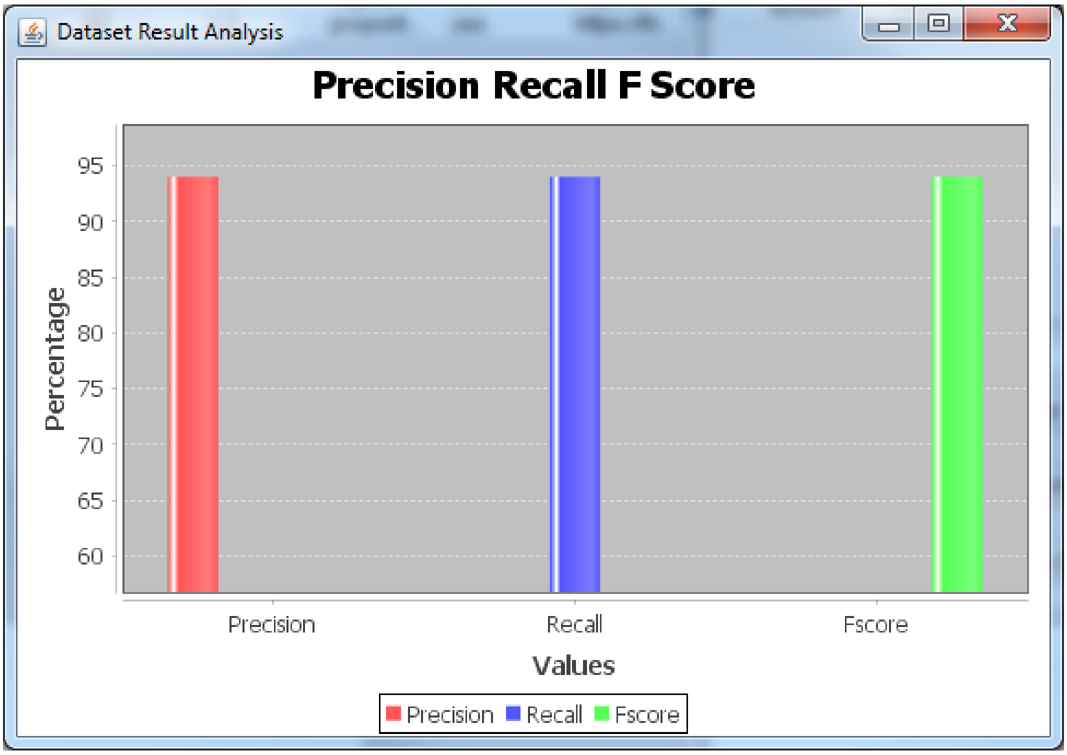

Figure 3.1 Precision Recall and FScore without Word2Vec

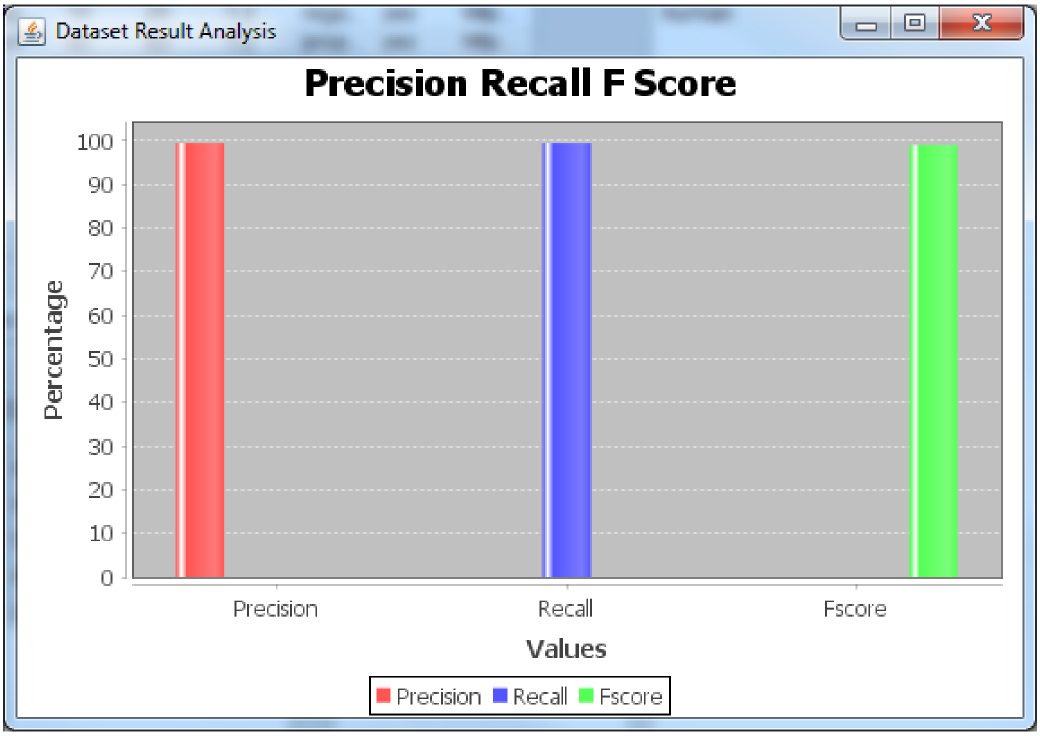

Figure 3.1 Precision Recall and FScore With Word2Vec

The most important category measurements for binary categories are:

Precision	Recall	F Measure
$P = TP/(TP + FP)$	$R = TP/(TP + FN)$	$tp + tn/tp + tn + fp + fn$

Table 3.1 Classification Results

Classification	Precision	Recall	FScore
Our Implementation	99.78	98.23	98.10
Authors	95.78	95.23	94.90

Conclusion

With better datasets and knowledge base, better results will be showcased by the bot. But since there are some major components where there have been dynamic changes. The IR (information retrieval) process tends to change with time. Different algorithms have been used and with technology evolving the retrieval process including the triggering part is getting faster. Not just IR but also the answering part with the word embedding technique coming into play makes it easier to relate or connect to different words. Such word embedding makes LSTM network more efficient. The tendencies of the results are promising in that a NLP and self-learning techniques would be a good addition to a chatbot as to make it more human-like. Although, further research would be required to make any real conclusions.