ISSN No:-2456-2165

A Comparative Study on Chatbot Based on Machine Learning and Lexicon Based Technique

Karthik Konar.
MCA Student,Dept of Computer Engineering,
NMIMS Mukesh Patel School Of Technology Management &
Engineering, Vile Parle(West) Mumbai.

Abstract:- Sentimental Analysis is that particular domain ,where you try to understand human emotions with the help of a software.Human emotions are in written form and we can classify those sentiments as positive,negative and neutral.Sentimental analysis is also referred to as opinion mining because in sentimental analysis we are trying to analyze the thoughts of a customer with respect to a particular thing.

However Natural Language Processing and Machine learning are considered to be the childrens of Artificial Intelligence,Since they both work in conjunction and lend a hand to solve large numbers of data problems.While Natural Language Processing provides us with an understanding about how computers and human(natural) language interact with each other.

This paper aims to identify which approach(lexicon or machine learning)is better among the two approaches in terms of providing accurate results when it is implemented in ChatBot.

Python language is utilized for the development of the chatbots.one chatbot is developed for classifying movie reviews as positive,negative or neutral by taking the input from the user and another chatbot(DocBot) is developed for providing all the information related to kidney disease to the user.

Keyword:- *Chatbot, Lexicon, Machine Learning, polarity, subjectivity, tokenization.*

I. INTRODUCTION

Nowadays customers play a very big role in making a business or any entity successful.A customer can make or break a business,therefore it is very important for the organization to understand the sentiments of its customers ,client's so that any organization can reach heights.Therefore Sentimental analysis is essential.Sentimental analysis determine useful information,those information can be used to understand current market strategy,improve business.There are various applications of Sentimental analysis such as Review classification,Product Review Mining.[1]States that sentimental analysis is is a system or a model that takes the document that analyzed the input ,and generates a detailed document summarizing the options of the given input document.

This paper provides the detailed comparison between lexicon based approach and machine learning based approach .[4]Chatbot refers to a chatting robot.[4] It is a communication simulating computer program. [4]It is all about the conversation with the user. [4]The conversation with a Chatbot is very simple. [4]It answers the questions asked by the user.[6]A chatbot, also known as a conversational agent, is a computer software capable of taking a natural language input and providing a conversational output in real time.[7]A chatbot is the best tool which provides a quick way to interact with the users. [7]It is very helpful to the users as it allows them to enter questions in natural language and desired information is obtained easily to the user. 2 chatbots(CHATBOT1,DOCBOT) were developed using the above mentioned approaches,and the chatbot which gives us the most accurate results are discussed in this paper along with advantage and disadvantage of each approach is also discussed in this paper. Chatbot1 which we have developed comes under lexicon based approach ,and that chatbot takes a list of words as an input from the user and it then identifies the polarity of the text.The main work of CHATBOT1 is to take reviews of movies from the user and then classify those reviews as positive,negative or neutral.This chatbot uses TEXTBLOB library for processing textual data.The concept of polarity and subjectivity is used while developing this chatbot(CHATBOT1).

DOCBOT is another chatbot which is developed using Machine Learning Based Approach.The DOCBOT provides information related to kidney disease to the user.The concept of tokenization(ie Lemmatization),TFIDVectorizer is used while developing DOCBOT.

This study helps us to compare which approach is better and provides us with the suitable results.

II. MATERIALS AND METHODS

- *Literature Survey*

There are 2 approaches which are extensively used to detect sentiments from the text.They are symbolic techniques and machine learning techniques.

[1]In their research work concluded that machine learning technique is very easier and efficient than symbolic techniques(Lexicon approach)

[2]Developed A Wat son chatbot which shows us and performs the tasks like "on headlamps" or "Turn on

ISSN No:-2456-2165

wipers".A user may input commands while driving through voice assist easily without any distraction from the road and The bot will perform those tasks for him.

[3]Introduced a new method method called B-Point Tree to speed up the search process by adding an additional data structure that contains shortcut pointers to the traditional search BST. The experiments had been conducted on a FloristBot, a chatbot that behaves as human personnel in a flower shop. The FloristBot is used to entertain customers and take orders.

[4] states that chatbot is one of the simple ways to transport data from a computer without having to think for proper keywords to look up in a search or browse several web pages to collect information.In her review paper she concluded that the development and improvement of chatbot design grow at an unpredictable rate due to variety of methods and approaches used to design a chatbot.

[5]in their paper stated that larger lexicons may yield a decrease in performance due to ambiguity of words polarity and increased model complexity.

[7]Developed a Chatbot which provides various information related to university or college and also students-related information. The chatbot can be used by anyone who can access the university's website. The project uses the concept of Artificial Intelligence and Machine Learning.

[8]Introduced the concept of CyberBullying in two way chat using machine learning algorithms,their main aim was to detect cyberbully in chatbot using cyberbully algorithm.

[9]Their paper explains a medical chatbot which can be used to replace the conventional method of disease diagnosis and treatment recommendation using machine learning approach.

[10] In their survey, the results showed that the greatest advantage of using chatbots in marketing is the provision of simple, fast information, but they also showed the fear of respondents getting the wrong information from chatbots, which is something that needs to be resolved in the future.

III. RESULTS AND DISCUSSION

➢ *System Architecture Of ChatBot1 Developed using Lexicon based approach*

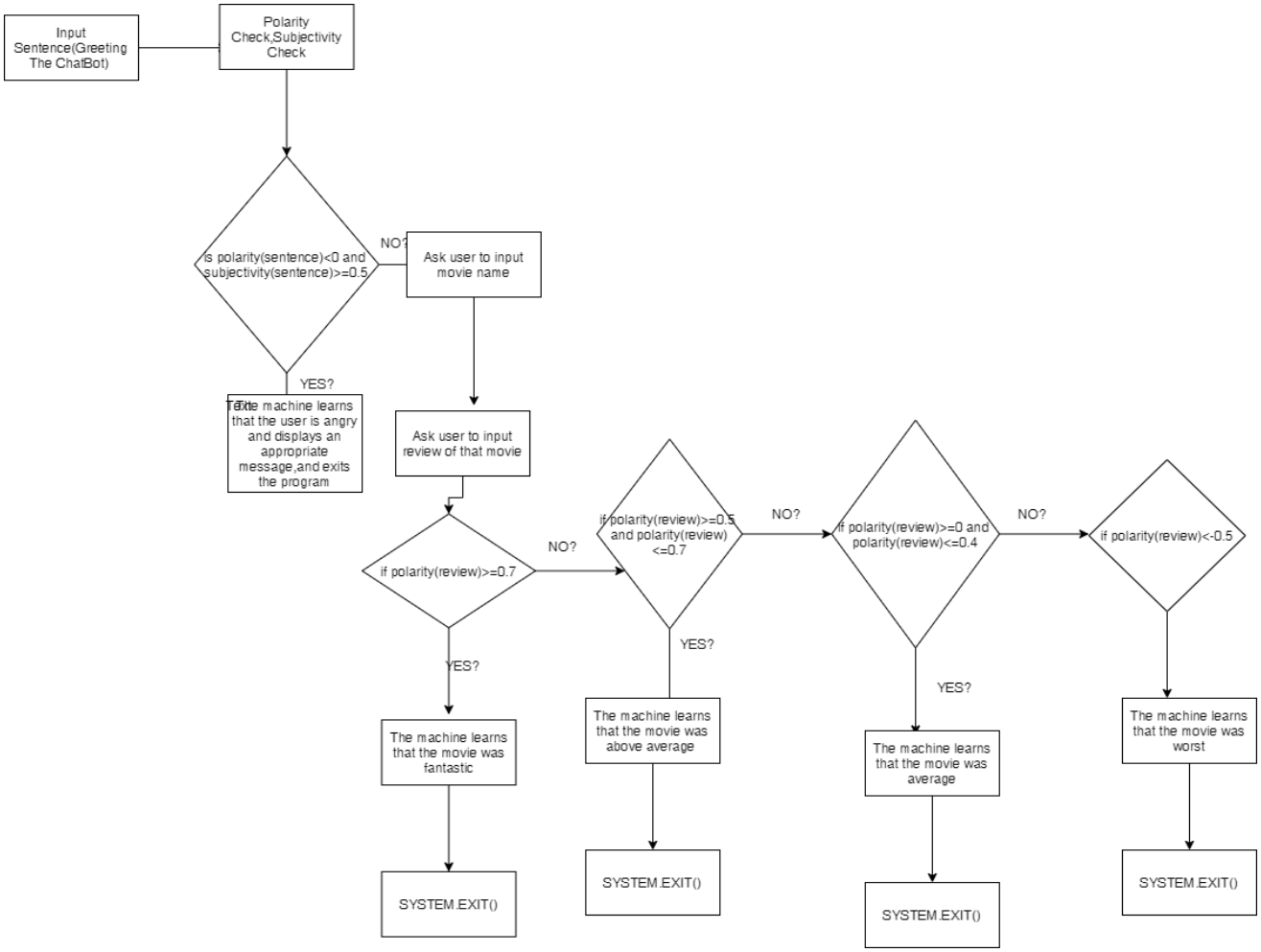

Fig 1:- System Architecture of ChatBot for Movie Review Classification Using Lexicon Approach.

The algorithm for the above chatbot is shown as follows:

Step 1: A greeting message is displayed to the user .The user is asked to greet back the chatbot . The greeting sentence of the user is then subjected to undergo polarity and subjectivity check.if the polarity of the sentence is less than 0 and the subjectivity of the sentence is greater than or equal to 0.5 the chatbot assumes that the user is angry and displays an appropriate message to the user and the chatbot terminates.

Step 2: if the polarity of the sentence is not less than 0 and the subjectivity of the sentence is less than 0.5 the chatbot assumes that the user is fine and happy and and the chatbot asks the user to input the movie name recently watched by him/her ,and then the chatbot asks the user to write a review about that particular movie.

The review is then subjected to measure polarity on the text.

If the polarity of the sentence:

- Is greater than or equal to 0.7,the chatbot assumes that the movie was fantastic and then it terminates.
- Is greater than or equal to 0.5 and less than 0.7 ,the chatbot assumes that the movie was above average and then it terminates.
- Is greater than or equal to 0 and less than 0.4 ,the chatbot assumes that the movie was average and then it terminates.
- Is less than -0.5 ,the chatbot assumes that the movie was worse.

This approach is known as the Lexicon based approach.

➢ *System Architecture of DocBot Developed using Machine Learning Technique.*

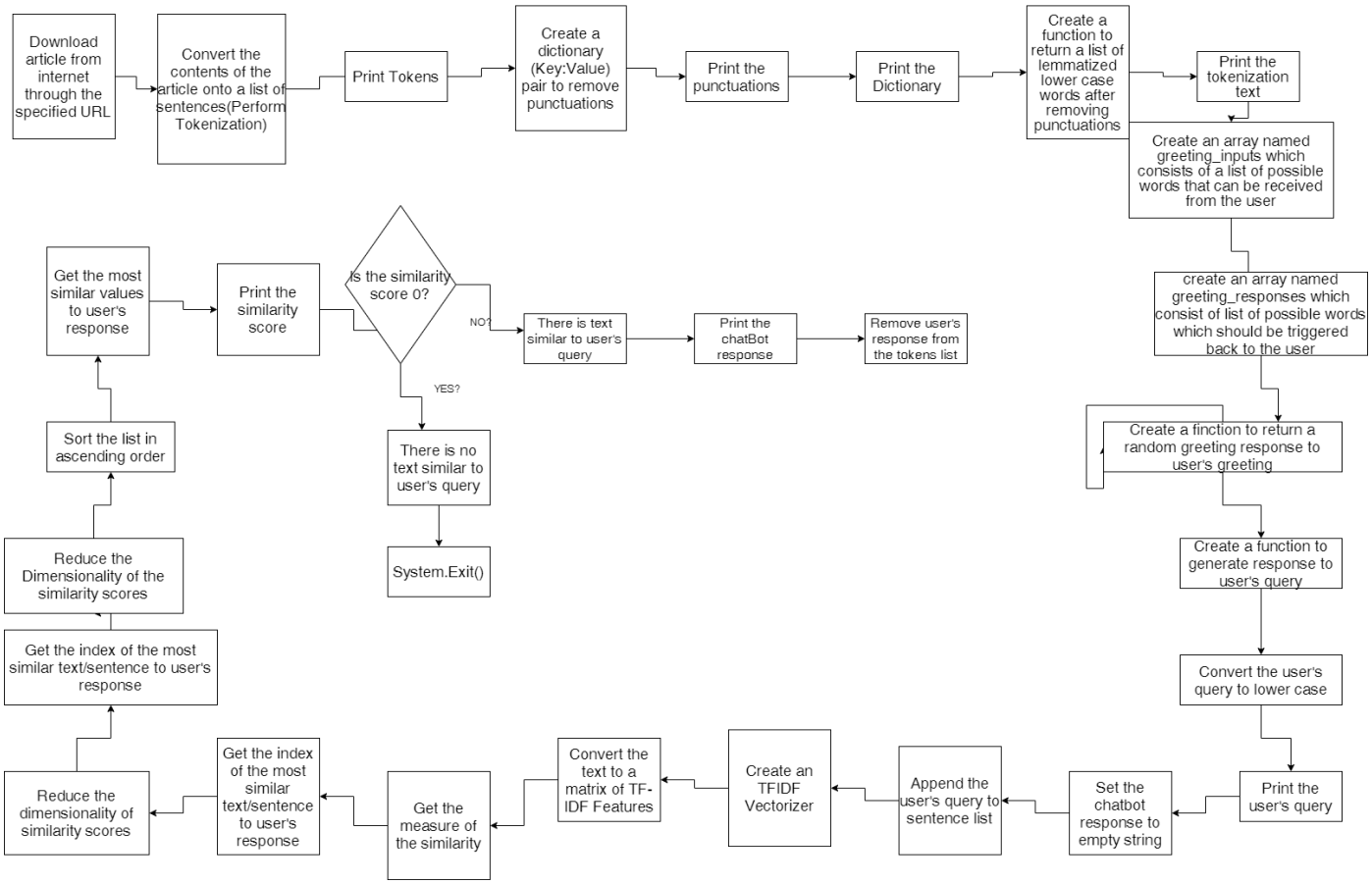

Fig 2:- System Architecture of ChatBot for Movie Review Classification Using Machine Learning Approach.

The algorithm for the above chatbot is shown as follows:

Step 1:Download article from the internet
pseudocode:article1=Article('SPECIFY URL')
article1.download()

```
)verview

Chronic kidney disease, also called chronic kidney failure, describes the gradual loss of kidney function. Your kidneys filter wastes and excess fluids from your blo
  which are then excreted in your urine. When chronic kidney disease reaches an advanced stage, dangerous levels of fluid, electrolytes and wastes can build up in yo
)ody.

In the early stages of chronic kidney disease, you may have few signs or symptoms. Chronic kidney disease may not become apparent until your kidney function is signi
:antly impaired.

Treatment for chronic kidney disease focuses on slowing the progression of the kidney damage, usually by controlling the underlying cause. Chronic kidney disease can
:ogress to end-stage kidney failure, which is fatal without artificial filtering (dialysis) or a kidney transplant.

Chronic kidney disease care at Mayo Clinic

How kidneys work

Symptoms

Signs and symptoms of chronic kidney disease develop over time if kidney damage progresses slowly. Signs and symptoms of kidney disease may include:

Nausea

Vomiting

Loss of appetite

Fatigue and weakness
```

It will download the entire article from the specified URL

Step 2: Convert the text into a list of sentences. Ie,perform tokenization.

Step 3:print tokens.

```
['Overview\n\nChronic kidney disease, also called chronic kidney failure, describes the gradual loss of kidney function.', 'Your kidneys filter wastes and excess fluic
s from your blood, which are then excreted in your urine.', 'When chronic kidney disease reaches an advanced stage, dangerous levels of fluid, electrolytes and wastes
can build up in your body.', 'In the early stages of chronic kidney disease, you may have few signs or symptoms.', 'Chronic kidney disease may not become apparent unti
l your kidney function is significantly impaired.', 'Treatment for chronic kidney disease focuses on slowing the progression of the kidney damage, usually by controlli
ng the underlying cause.', 'Chronic kidney disease can progress to end-stage kidney failure, which is fatal without artificial filtering (dialysis) or a kidney transpl
ant.', 'Chronic kidney disease care at Mayo Clinic\n\nHow kidneys work\n\nSymptoms\n\nSigns and symptoms of chronic kidney disease develop over time if kidney damage p
rogresses slowly.', "Signs and symptoms of kidney disease may include:\n\nNausea\n\nVomiting\n\nLoss of appetite\n\nFatigue and weakness\n\nSleep problems\n\nChanges i
n how much you urinate\n\nDecreased mental sharpness\n\nMuscle twitches and cramps\n\nSwelling of feet and ankles\n\nPersistent itching\n\nChest pain, if fluid builds
up around the lining of the heart\n\nShortness of breath, if fluid builds up in the lungs\n\nHigh blood pressure (hypertension) that's difficult to control\n\nSigns ar
d symptoms of kidney disease are often nonspecific, meaning they can also be caused by other illnesses.", 'Because your kidneys are highly adaptable and able to compen
sate for lost function, signs and symptoms may not appear until irreversible damage has occurred.', 'When to see a doctor\n\nMake an appointment with your doctor if yo
u have any signs or symptoms of kidney disease.', 'If you have a medical condition that increases your risk of kidney disease, your doctor is likely to monitor your bl
ood pressure and kidney function with urine and blood tests during regular office visits.', 'Ask your doctor whether these tests are necessary for you.', 'Causes\n\nNo
rmal kidney vs. diseased kidney Open pop-up dialog box Close Normal kidney vs. diseased kidney Normal kidney vs. diseased kidney A normal kidney has about 1 million fi
ltering units.', 'Each unit, called a glomerulus, connects to a tubule, which collects urine.', 'Conditions such as high blood pressure and diabetes take a toll on kid
```

Step 4:create a dictionary (key:value) pair to remove the punctuations.
Step 5: print the punctuations.

```
!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
```

Step 6:Print the dictionary

```
{33: None, 34: None, 35: None, 36: None, 37: None, 38: None, 39: None, 40: None, 41: None, 42: None, 43: None, 44: None, 45: None, 46: None, 47: None, 58: None, 59: No
ne, 60: None, 61: None, 62: None, 63: None, 64: None, 91: None, 92: None, 93: None, 94: None, 95: None, 96: None, 123: None, 124: None, 125: None, 126: None}
```

Step 7:create a function to return a list of lemmatized lower case words after removing punctuations.

PseudoCode:
def LemNormalize1(text1):
 return nltk.word_tokenize(text1.lower().translate(remove_punc_dict))
The above is created to remove all the punctuations from the article.

Step 8: printing the tokenization text.

```
['overview', 'chronic', 'kidney', 'disease', 'also', 'called', 'chronic', 'kidney', 'failure', 'describes', 'the', 'gradual', 'loss', 'of', 'kidney', 'function', 'your
', 'kidneys', 'filter', 'wastes', 'and', 'excess', 'fluids', 'from', 'your', 'blood', 'which', 'are', 'then', 'excreted', 'in', 'your', 'urine', 'when', 'chronic', 'ki
dney', 'disease', 'reaches', 'an', 'advanced', 'stage', 'dangerous', 'levels', 'of', 'fluid', 'electrolytes', 'and', 'wastes', 'can', 'build', 'up', 'in', 'your', 'bod
y', 'in', 'the', 'early', 'stages', 'of', 'chronic', 'kidney', 'disease', 'you', 'may', 'have', 'few', 'signs', 'or', 'symptoms', 'chronic', 'kidney', 'disease', 'may'
, 'not', 'become', 'apparent', 'until', 'your', 'kidney', 'function', 'is', 'significantly', 'impaired', 'treatment', 'for', 'chronic', 'kidney', 'disease', 'focuses',
'on', 'slowing', 'the', 'progression', 'of', 'the', 'kidney', 'damage', 'usually', 'by', 'controlling', 'the', 'underlying', 'cause', 'chronic', 'kidney', 'disease', '
can', 'progress', 'to', 'endstage', 'kidney', 'failure', 'which', 'is', 'fatal', 'without', 'artificial', 'filtering', 'dialysis', 'or', 'a', 'kidney', 'transplant', '
chronic', 'kidney', 'disease', 'care', 'at', 'mayo', 'clinic', 'how', 'kidneys', 'work', 'symptoms', 'signs', 'and', 'symptoms', 'of', 'chronic', 'kidney', 'disease',
'develop', 'over', 'time', 'if', 'kidney', 'damage', 'progresses', 'slowly', 'signs', 'and', 'symptoms', 'of', 'kidney', 'disease', 'may', 'include', 'nausea', 'vomiti
ng', 'loss', 'of', 'appetite', 'fatigue', 'and', 'weakness', 'sleep', 'problems', 'changes', 'in', 'how', 'much', 'you', 'urinate', 'decreased', 'mental', 'sharpness',
'muscle', 'twitches', 'and', 'cramps', 'swelling', 'of', 'feet', 'and', 'ankles', 'persistent', 'itching', 'chest', 'pain', 'if', 'fluid', 'builds', 'up', 'around', 't
he', 'lining', 'of', 'the', 'heart', 'shortness', 'of', 'breath', 'if', 'fluid', 'builds', 'up', 'in', 'the', 'lungs', 'high', 'blood', 'pressure', 'hypertension', 'th
ats', 'difficult', 'to', 'control', 'signs', 'and', 'symptoms', 'of', 'kidney', 'disease', 'are', 'often', 'nonspecific', 'meaning', 'they', 'can', 'also', 'be', 'caus
ed', 'by', 'other', 'illnesses', 'because', 'your', 'kidneys', 'are', 'highly', 'adaptable', 'and', 'able', 'to', 'compensate', 'for', 'lost', 'function', 'signs', 'an
d', 'symptoms', 'may', 'not', 'appear', 'until', 'irreversible', 'damage', 'has', 'occurred', 'when', 'to', 'see', 'a', 'doctor', 'make', 'an', 'appointment', 'with',
'your', 'doctor', 'if', 'you', 'have', 'any', 'signs', 'or', 'symptoms', 'of', 'kidney', 'disease', 'if', 'you', 'have', 'a', 'medical', 'condition', 'that', 'increase
s', 'your', 'risk', 'of', 'kidney', 'disease', 'your', 'doctor', 'is', 'likely', 'to', 'monitor', 'your', 'blood', 'pressure', 'and', 'kidney', 'function', 'with', 'ur
ine', 'and', 'blood', 'tests', 'during', 'regular', 'office', 'visits', 'ask', 'your', 'doctor', 'whether', 'these', 'tests', 'are', 'necessary', 'for', 'you', 'causes
', 'normal', 'kidney', 'vs', 'diseased', 'kidney', 'open', 'popup', 'dialog', 'box', 'close', 'normal', 'kidney', 'vs', 'diseased', 'kidney', 'normal', 'kidney', 'vs',
```

Step 9:create an array named GREETING_INPUTS1 which contains a list of words which can be received as a greeting message from the user.

pseudocode:GREETING_INPUTS1=["hi","hello","hola","wassup","hey"]

Step 10:create an array named GREETING_RESPONSES1 which should be triggered back to the user.
pseudocode:GREETING_RESPONSES=["howdy","hi","hey","what's good","hello","hey there"]

Step 11: create Function to return a random greeting response to a user's greeting.

Step 12: create a function to generate the response to user's query

Step 13:convert the user's query to lower case.

Step 14:print the user query

Step 15:set the chatbot response to an empty string.

Step 16:Append the user's query to sentence list

Step 17:Create a TFIDF Vectorizer and print it's features.

```
(0, 104)      0.21517271105967337
(0, 142)      0.31781973567163924
(0, 109)      0.36025299997679194
(0, 64)       0.36025299997679194
(0, 90)       0.31781973567163924
(0, 31)       0.2643601450098346
(0, 72)       0.1426325666011141
(0, 133)      0.37394360688199496
(0, 45)       0.3636909501497882
(0, 173)      0.36025299997679194
(1, 247)      0.29621491649318066
(1, 88)       0.40366263341462977
(1, 22)       0.24110051322517356
(1, 101)      0.40366263341462977
(1, 87)       0.40366263341462977
(1, 257)      0.35611626124035106
(1, 97)       0.40366263341462977
(1, 134)      0.274835201145623
(2, 23)       0.28616745795373816
(2, 29)       0.32437471199107215
(2, 80)       0.32437471199107215
(2, 99)       0.2590589798027162
(2, 138)      0.28616745795373816
(2, 62)       0.32437471199107215
(2, 221)      0.32437471199107215
:     :
(20, 206)     0.16848899614087914
(20, 131)     0.0930600200686559
(20, 185)     0.07181971765373221
(20, 113)     0.07740664797917175
(20, 144)     0.0930600200686559
(20, 111)     0.16848899614087914
(20, 239)     0.0930600200686559
```

Step 18:convert the text of a matrix to TF IDF Features.Get the measure of the similarity scores from the user query.
Using cosine_similarity module.

ISSN No:-2456-2165

```
(21, 133)     0.65803916079193B9
[0.35086378 0.         0.17030105 0.28157416 0.36438763 0.25087682
 0.34328057 0.34089121 0.15195873 0.         0.21918276 0.20807616
 0.         0.24042557 0.         0.0715544  0.20025711 0.20699116
 0.42121216 0.36152496 0.16639361 1.        ]]
```

Step 19:Get the index of the most similar text/sentence to the users response

Step 20:sort the list in ascending order.

Step 21:Get the most similar score to the users response

Step 22: Print the similarity score

Step 23:if the similarity score is 0 then there is no text similar to the users query.

Step 24:if the similarity score is non-zero print the chatbot response and user's query from the user's token list.

- *Results*

The following results were obtained for the chatbot which was implemented using Lexicon Technique.

```
Hello!! My name Is ChatBot.Press Enter To Continue

Hey there.I oouldn't recognize you.you are...?
you are the worst chatbot ever
hey..you need to calm down..i just asked your name
come back when you are calm..byee
```

Fig 3:- In the above figure the polarity measured from the Greeting sentence is less than 0 .So therefore the chatbot learns that the user is angry,therefore displays appropriate messages and then terminates .

```
Hello!! My name Is ChatBot.Press Enter To Continue

Hello there.May I know your name
My name is karthik
are you sure?
yes
Please input your name again to continue
karthik
karthik..Gotcha..!!
All Right.Now what's your gender.Male or Female?
Male
ohkk......so let's start talking
What was the last movie you watched?
student of the year
Oh wow.. I want to watch that movie too..!! How is it?
the movie was awful
Oh No...Was it that bad? I hope the tickets get refunded
```

Fig 4:- In the above figure the polarity measured from the sentence is less than -0.5 .so therefore the chatbot learns that the movie is worse.

```
What was the last movie you watched?
kota factory
Oh wow.. I want to watch that movie too..!! How is it?
the movie was wonderful with an amazing experience
I knew it would be fantastic..Gotta watch it soon
```

Fig 5:- In the above figure the polarity measured from the sentence is greater than 0.7 .so therefore the chatbot learns that the movie is wonderful.

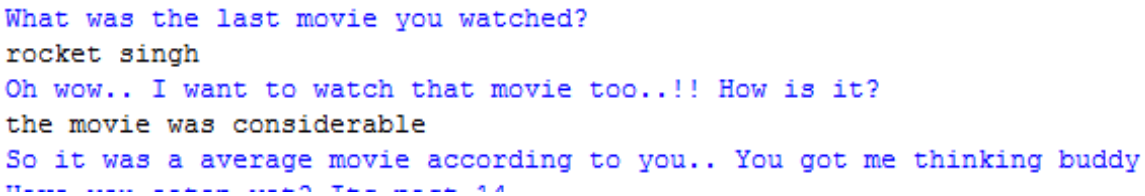

Fig 6:- In the above figure the polarity measured from the sentence is greater than 0 and less than or equal to 0.4 .so therefore the chatbot learns that the movie is average.

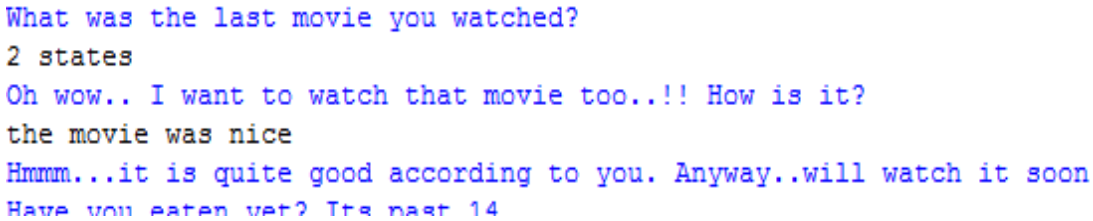

Fig 7:- In the above figure the polarity measured from the sentence is greater than or equal to 0.5 and less than or equal to 0.7 .so therefore the chatbot learns that the movie is above average.

The following results were obtained for the chatbot which was implemented using Machine Learning Technique.

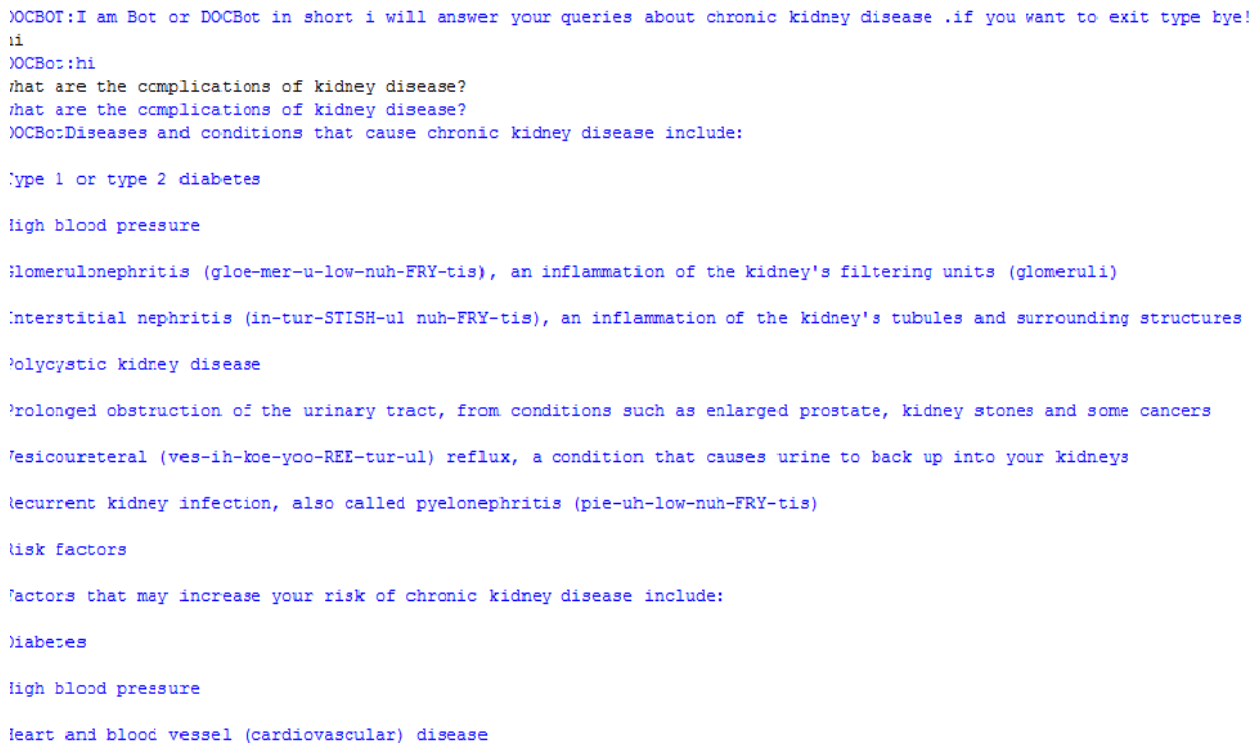

Fig 8:- DocBot responding to a user's query.

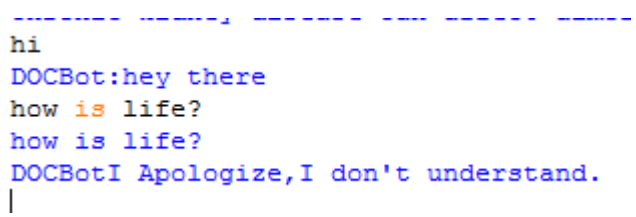

Fig 9:- if there is no similarity found in the user's query then the chatbot prints appropriate messages.

➢ *Findings:*
Based on the experimentation performed, the following findings were obtained:

Sr. No	Technique	Advantages	Disadvantage	Remarks.
1	Lexicon	Easy to implement,Easy to understand ,Less complex when compared with machine learning approach.	Accuracy rate is low when compared with machine learning approach,Based on WordNet Database.	It is implemented on chatbot which classifies movies as wonderful,above average,average,worst based on user reviews
2	Machine Learning	Accuracy rate is very much higher when compared with Lexicon approach,Very good performance.	Complex to implement when compared with lexicon approach	It is implemented on chatbot which answers all queries related to kidney disease.

Table 1:- Summary of Comparison.

IV. CONCLUSION

➢ *2 chatbots were developed*
First chatbot is developed for classifying movie reviews as positive,negative or neutral by taking the input from the user and another chatbot(DocBot) is developed for providing all the information related to kidney disease to the user.When compared the performace and efficiency of both the chatbots it was observed that the chatbot which was developed using machine learning approach proved to produce more promosing and faster results than the chatbot which was developed using lexicon approach.

Thus I conclude that machine learning techniques are more efficient than Lexicon based approaches when it needs to be implemented in Chatbot.

V. FUTURE SCOPE

This paper provides a detailed study on why machine learning approach is better than lexicon approach while implementing chatbot.Hence this paper will prove to be useful for upcoming authors who wish to make a further detailed analysis between machine learning approach and lexicon based approach.

ACKNOWLEDGMENT

I acknowledge the contribution of NMIMS university to provide me with this amazing opportunity and good facilities to carry out this review work.

REFERENCES

[1]. Akshay Amolik,Niketan Jivane,Mahavir Bhandari,Dr.M Venkatesan.'Twitter Sentiment Analysis of Movie Reviews Using Machine Learning Techniques'.International Journal of Engineering and Technology.
[2]. Praveen Kumar,Mayank Sharma,Seema Rawat,Tanupriya Choudhury.'Designing and Developing a ChatBot using Machine Learning'.Proceedings of the SMART–2018, IEEE Conference ID: 44078 2018 International Conference on System Modeling & Advancement in Research Trends, 23rd–24th November, 2018 College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India
[3]. Ayah Atiyah ,Shaidah Jusoh:Sufyan Almajali.'An Efficient Search for Context-Based Chatbots'2018 8th International Conference on Computer Science and Information Technology (CSIT).
[4]. M. Dahiya.'A Tool of Conversation: Chatbot'.International Journal of Computer Sciences and Engineering .Volume-5, Issue-5. E-ISSN: 2347-2693.
[5]. Olga Kolchyna , Tharsis T. P. Souza ´ , Philip C. Treleaven and Tomaso Aste.'Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination'.
[6]. Prissadang Suta,Xi Lan , Biting Wu,Pornchai Mongkolnam and Jonathan H. Chan.'An Overview of Machine Learning in Chatbots ' International Journal of Mechanical Engineering and Robotics Research Vol. 9, No. 4, April 2020.
[7]. Neelkumar P. Patel, Devangi R. Parikh,Prof. Darshan A. Patel,Prof. Ronak R. Patel,Prof. Darshan A. Patel3 Prof. Ronak R. Patel.'AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT) 'Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019] IEEE Conference Record # 45616; IEEE Xplore ISBN: 978-1-7281-0167-5
[8]. Mrs.V.Selvi, Ms.Saranya , Ms.Chidida , Ms.Abarna.'Chatbot and bullyfree Chat.'Proceding of International Conference on Systems Computation Automation And Networking 2019.IEEE 978-1-7281-1524-5.
[9]. Rohit Binu Mathew, Sandra Varghese, Sera Elsa Joy, Swanthana Susan Alex.'Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning'.Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.
[10]. uros arsenijevic,marija jovic 'Artificial intelligence marketing: Chatbots'.2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI) .