

ChatBot which Identifies Depression

Rohit Gollarahalli (B00779758)

MACS, Natural Language Processing, CSCI 6509
Dalhousie University
Halifax, Canada
rh522064@dal.ca

Nibir Mukherjee (B00768437)

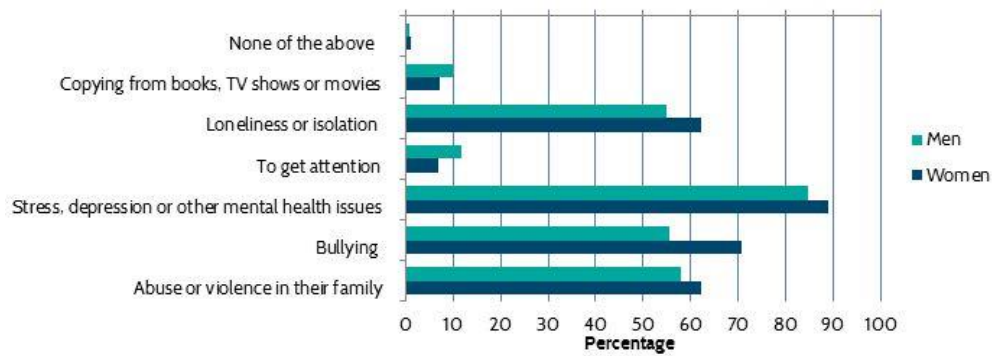
MACS, Natural Language Processing, CSCI 6509
Dalhousie University
Halifax, Canada
nb75571@dal.ca

Abstract – This project was started with a motive to solve one of the most significant problem of today's world, i.e. depression. With rising technologies, challenges, demands and needs, today's generation often find themselves in a situational deadlock and start building stress within themselves. This stress results to depression and depression results in suicide. Statistics say that over 50 percent of the people who die by suicide, suffer from depression. So, in order to fix this issue or rather try to lower down the statistics, we aimed to build a program/bot which will act as a therapist, engaging itself in a conversation with a person and then, on the basis of the conversation, will flag the person as depressed or non-depressed. This chatbot initiates a conversation and works on pattern matching. When the user decides to quit talking, the model will send the stored text, which the user inputs as a part of the conversation, to a classifier for our supervised learning. We did a comparative study for different machine learning models by applying different algorithms, and taking into consideration the unigram or bigram model, to identify the one with the best performance.

I. INTRODUCTION

Today, in 21st century, the world has seen an advent of what is called a technological bloom. In this advanced era, human beings have found several ways to prove them as an individual in different grounds and aspects. Illiteracy rate is decreasing, and lifestyle of people is improving day by day. So, with all these improvements going arounds, we assume that we are heading towards a 'happy world, happy life' era. But, when we look at the headlines, browse through the internet, we see that we are not even close to our assumptions. In spite of all the advancements, enhanced lifestyle, supreme knowledge, sheer happiness is still something which today's world hasn't provided us yet. With everything and everyone around us thriving towards achieving more, getting better every single day, we sometimes end up taking a heavy toll on ourselves. Life was short and simple in our previous generations. Increased competition to stand out amongst others, pursuing higher standards of life, increased demands, have all led to several complexities of life. And, all these complexities have led to the cause of one of the rising threat of human life, i.e. depression. Instead of achieving every materialistic pleasure of life, we still find people giving up their life out of depression. These people are scared, lonely and unable to speak about their problems with anyone else. Compared to other doctor specializations, we find psychotherapists to be quite less in numbers. So, this adds to the severity of the problem. Over internet, we see several blogs, with anonymous users suffering from depression, speaking about their lives and problems. Hardly do they get the help that they need and end up giving up their lives. But, in addition to that, there remains a huge number of people who are even hesitant to speak with anyone regarding their problems. So, with all these cases around we see the count of suicides increasing at an exponential rate every year. According to a study by a company named Relationship Australia, almost 90% of the suicide cases of today are due to stress and depression^[1].

Figure 4. Reasons for suicide risk, subjective reports
Monthly online survey, July 2017, final survey results



[Source: <https://www.relationships.org.au/what-we-do/research/online-survey/july-2017-youth-suicide/>]

Concerned by the rising threat and in order to lend a helping hand to solve the issue, we aimed to develop a chatbot which will act as a therapist in identifying depression. There were two prime motives in building so. Firstly, this will bridge the gap between the number of patients suffering from depression and the limited number of psychotherapists available to help the patients. And secondly, today's generation are more inclined towards textual conversation rather than speaking to someone in person. So for people who fail to express their problems to anyone else, can take the help of this bot therapist to share their problems. The bot aims to solve two purposes; engage in conversation with a depressed person to try to help him, and secondly, gather information from the user to classify him as depressed or not if he is unsure of his current mental status. In this age of automation, this product can also be used by the therapists to automate their therapy sessions. There is no pre-defined process to bring out a person from depression, but studies say, that conversation is the best way to communicate distress and receive emotional and mental support.

II. RELATED WORK

There are a handful of efforts which have been taken in the past to solve the problem of depression. In order to build our therapy bot, we went through some research papers and products which dealt in the similar field. In 1966, a professor named Joseph Weizenbaum developed a program named ELIZA which was a natural language processing computer program which aimed to mock a conversation between a user and a psychotherapist^[2]. Later in 2016, three students from John Hopkins University along with a professor from University of Tulsa, submitted a paper on 'Identifying Depression on Twitter' which proposes a model that will identify depressed persons from their activities over social media.

a) ELIZA

In 1966, neither technology nor complexities of human life had reached that extent as that of today. But researchers did feel the need of building a therapist which can help in treating depression patients because the issue of having less number of psychotherapist was prevalent in those times also. So, Joseph Weizenbaum developed a chatbot called Eliza. This chatbot works on pattern matching. Conditions are based on some regular expressions, which consists of sentences which can mock the therapy session of a real-world psychotherapist. For digging more into this chatbot, we went through the research paper that was written by Joseph Weizenbaum in 1966 on Computational Linguistics^[2]. Quoting from the paper, "The fundamental technical problems with which ELIZA is concerned are: (1) the identification of key words, (2) the discovery of minimal context, (3) the choice of appropriate transformations, (4) generation of responses in the absence of key words, and (5) the provision of an editing capability for ELIZA "scripts""^[2]. So this paper leaves some future aspects to be handled. Our project is inspired from the problems that Eliza was not able to handle. In addition to the above

mentioned drawbacks, Eliza lacks the potential to handle modern-world problems as the pattern matching conditions which are implemented works in general with the inputs entered by the user and not extend on some particular case-specific scenarios. Moreover, the chatbot provides only therapy, lacking the ability to conclude anything about the user. So, in our project, we tried to handle the drawbacks of Eliza and gave it an extension by programming a classifier, which can classify the inputs entered by an user and conclude on the patient, based on the inputs. The below screenshot shows a snapshot of a conversation with the Eliza bot.

```
Welcome to

      EEEEE LL      IIII ZZZZZZ  AAAA
      EE      LL      II      ZZ  AA  AA
      EEEEE LL      II      ZZ  AAAAAA
      EE      LL      II      ZZ  AA  AA
      EEEEE LLLLLL IIII ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

[Source: <http://www.le-grenier-informatique.fr/pages/les-documents-du-grenier/le-chatbot-eliza.html>]

The pattern matching script on which Eliza works is given at the bottom of the research paper published by Joseph Weizenbaum, in the Appendix section. A snapshot of the Appendix is shown below.

APPENDIX. An ELIZA Script

```
(HOW DO YOU DO. PLEASE TELL ME YOUR PROBLEM)
START
(SORRY ((0) (PLEASE DON'T APOLOGIZE)
(APOLOGIES ARE NOT NECESSARY) (WHAT FEELINGS
DO YOU HAVE WHEN YOU APOLOGIZE) (I'VE TOLD YOU
THAT APOLOGIES ARE NOT REQUIRED)))
(DONT = DON'T)
(CANT = CAN'T)
(WONT = WON'T)
(REMEMBER 5
((0 YOU REMEMBER 0) (DO YOU OFTEN THINK OF 4)
(DOES THINKING OF 4 BRING ANYTHING ELSE TO MIND)
(WHAT ELSE DO YOU REMEMBER)
(WHY DO YOU REMEMBER 4 JUST NOW)
(WHAT IN THE PRESENT SITUATION REMINDS YOU OF 4)
(WHAT IS THE CONNECTION BETWEEN ME AND 4))
((0 DO I REMEMBER 0) (DID YOU THINK I WOULD FORGET 5)
(WHY DO YOU THINK I SHOULD RECALL 5 NOW)
(WHAT ABOUT 5) (=WHAT) (YOU MENTIONED 5))
((0) (NEWKEY)))
(IF 3 ((0 IF 0) (DO YOU THINK ITS LIKELY THAT 3) (DO YOU WISH THAT 3)
(WHAT DO YOU THINK ABOUT 3) (REALLY, 2 3)))
(WHAT DO YOU THINK ABOUT 3) (REALLY, 2 3)))
(DREAMT 4 ((0 YOU DREAMT 0)
(REALLY, 4) (HAVE YOU EVER FANTASIED 4 WHILE YOU WERE AWAKE)
(HAVE YOU DREAMT 4 BEFORE) (=DREAM) (NEWKEY)))
(DREAMED = DREAMT 4 (=DREAMT))
(DREAM 3 ((0) (WHAT DOES THAT DREAM SUGGEST TO YOU)
(DO YOU DREAM OFTEN) (WHAT PERSONS APPEAR IN YOUR DREAMS)

(WOULD YOU WANT TO BE 4) (YOU WISH I WOULD TELL YOU YOU ARE 4)
(WHAT WOULD IT MEAN IF YOU WERE 4) (=WHAT))
((0) (WHY DO YOU SAY 'AM') (I DON'T UNDERSTAND THAT)))
(ARE ((0 ARE I 0)
(WHY ARE YOU INTERESTED IN WHETHER I AM 4 OR NOT)
(WOULD YOU PREFER IF I WEREN'T 4) (PERHAPS I AM 4 IN YOUR
FANTASIES) (DO YOU SOMETIMES THINK I AM 4) (=WHAT))
((0 ARE 0) (DID YOU THINK THEY MIGHT NOT BE 3)
(WOULD YOU LIKE IT IF THEY WERE NOT 3) (WHAT IF THEY WERE NOT 3)
(POSSIBLY THEY ARE 3)) )
(YOUR = MY ((0 MY 0) (WHY ARE YOU CONCERNED OVER MY 3)
(WHAT ABOUT YOUR OWN 3) (ARE YOU WORRIED ABOUT SOMEONE ELSE'S 3)
(REALLY, MY 3)))
(WAS 2 ((0 WAS YOU 0)
(WHAT IF YOU WERE 4) (DO YOU THINK YOU WERE 4)
(WERE YOU 4) (WHAT WOULD IT MEAN IF YOU WERE 4)
(WHAT DOES ' 4 ' SUGGEST TO YOU) (=WHAT))
((0 YOU WAS 0)
(WERE YOU REALLY) (WHY DO YOU TELL ME YOU WERE 4 NOW)
(WERE YOU REALLY) (WHY DO YOU TELL ME YOU WERE 4 NOW)
(PERHAPS I ALREADY KNEW YOU WERE 4) )
((0 WAS I 0) (WOULD YOU LIKE TO BELIEVE I WAS 4)
(WHAT SUGGESTS THAT I WAS 4)
(WHAT DO YOU THINK) (PERHAPS I WAS 4)
(WHAT IF I HAD BEEN 4))
((0) (NEWKEY)) )
(WERE = WAS (=WAS))
(ME = YOU)
(YOU'RE = I'M ((0 I'M 0) (PRE (I ARE 3) (=YOU))))
(I'M = YOU'RE ((0 YOU'RE 0) (PRE (YOU ARE 3) (=I))))
(MYSELF = YOURSELF)
(YOURSELF = MYSELF)
(MOTHER DLIST(/NOUN FAMILY))
```

[Source: <https://dl.acm.org/citation.cfm?id=365168>]

b) Identifying Depression on Twitter

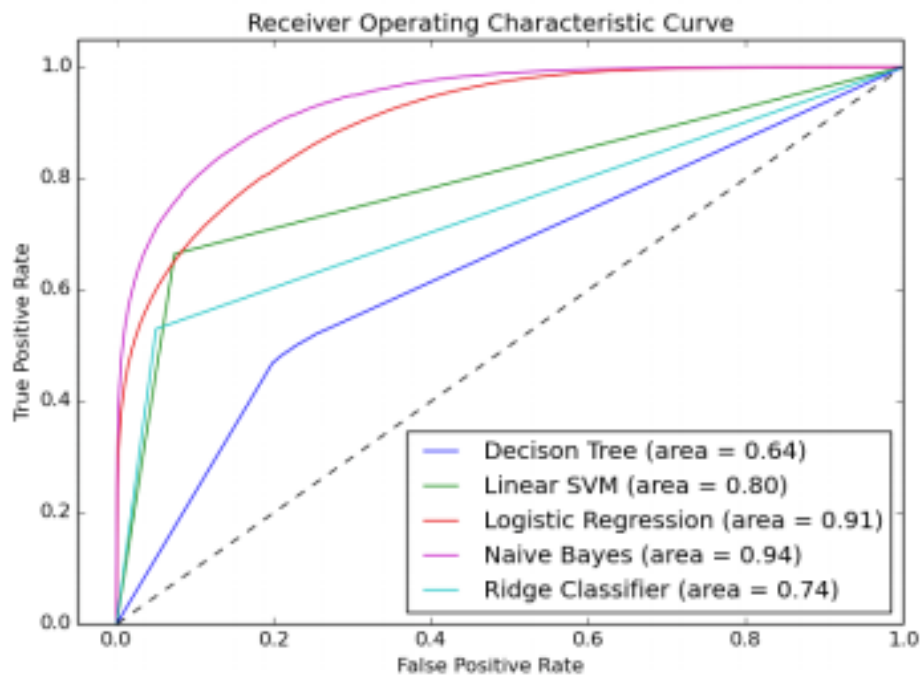
Mental health continues to be a growing concern amongst today's civilized world. It was reported that one in four Americans suffer from mental disorder. Recent statistics state that around 300 million people suffer from depression. Researchers recently discovered a correlation between depression diagnosis and social media. The study aimed at identifying depressed users by the means of twitter data. The current methods to identify depression have proven to be inefficient as only a certain percentage of today's government provide some form of primary care to combat mental illness while the rest do not provide any kind of institution for mental outreach^[3]. Such major depressive disorder when treated, tend to have positive impacts on the society. The dataset that was obtained was created by Coppersmith et al. for the Computational Linguistics and Clinical Psychology (CLPsych) 2015 Shared Task^[3]. These were twitter users who stated that they were diagnosed with depression. The study aimed at identifying, detecting and pursuing a diagnosis of individuals on twitter posts.

The definition of metrics used were Precision, Recall and F1 score. Classifiers were used to estimate the likelihood of depression in the user. For each classifier, Scikit-Learn from Pedregosa et al was utilized for implementation. The algorithms that were decided were Decision Tree, a Linear Support Vector classifier, Logistic Regression approach as well as Naïve Bayes. Through the approach of bag of words model an attempt to quantify depression from each word frequencies were carried out. The test results were as follows:

<i>Classification Algorithm</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>	<i>Samples</i>
<i>Decision Trees</i>	0.67	0.68	0.75	0.67	332421
<i>Linear Support Vector Classifier</i>	0.83	0.83	0.83	0.82	332421
<i>Naïve Bayes w/ 2-grams</i>	0.82	0.82	0.82	0.82	332421
<i>Logistic Regression</i>	0.86	0.82	0.84	0.82	332421
<i>Naïve Bayes w / 1-gram</i>	0.81	0.82	0.81	0.86	332421
<i>Ridge Classifier</i>	0.81	0.79	0.78	0.79	332421

[Source: <https://arxiv.org/ftp/arxiv/papers/1607/1607.07384.pdf>]

Naïve Bayes fell a little short of Logistic Regression in the classification task as it achieved a precision and F1 score of 0.81 while Logistic Regression achieved a precision of 0.86.



[Source: <https://arxiv.org/ftp/arxiv/papers/1607/1607.07384.pdf>]

In conclusion it was demonstrated how twitter can be utilized as a mean to identify depression.

III. PROBLEM DEFINITION AND METHODOLOGY

Our application is titled as the title of the paper states, “ChatBot which identifies depression”. After having gained the knowledge of the magnitude of the problem that we discussed, we decided to build an application which can lend a helping hand in solving this particular problem. We went through multiple research papers on this topic and have chosen the above two research papers, which serve as a base source, helping us in building our application.

For our application, we first checked the systems that have been built so far and few improvements that could have been made. In doing so, we found the drawbacks of chatbot Eliza and also noticed that there is no such application which can classify a user as depressed or not. Taking these areas into consideration, we planned to design a chatbot which will work like Eliza, but in an improved, enhanced way, trying to provide therapy sessions for modern-world problems. Through our supervised learning, we provided an extension to this. We store the conversation that takes place between the user and the chatbot and feed it to a classifier program. The classifier uses different machine learning algorithms, which run on the conversation text provided to it. After running successfully, it provides an output, classifying the user as depressed or not, along with the accuracy percentage of the classification.

There are two parts of our project. The first part being the chatbot, and the second being the classifier. So, for the first part, we looked on the existing code of Eliza. Eliza works on basic pattern matching algorithm. It checks the input chat entered by the user and matches it with the rules based on some regular expressions. Our chatbot also works in the similar way. We have written the code in python. Firstly, we import the required packages: re (for regular expressions), random (generating pseudo-random numbers) and reflections from nltk.chat.util (for reflecting words, i.e. changing you to I and vice-versa). The below image shows a screenshot of how reflections list looks like.

```
reflections = {
    "i am"      : "you are",
    "i was"     : "you were",
    "i"         : "you",
    "i'm"       : "you are",
    "i'd"       : "you would",
    "i've"      : "you have",
    "i'll"      : "you will",
    "my"        : "your",
    "you are"   : "I am",
    "you were"  : "I was",
    "you've"    : "I have",
    "you'll"    : "I will",
    "your"      : "my",
    "yours"     : "mine",
    "you"       : "me",
    "me"        : "you"
}
```

[Source: <http://www.nltk.org/modules/nltk/chat/util.html>]

Then comes the list of conditions based on a regular expression. We have taken a set of conditions that was already there in Eliza, as these conditions were written with the help of a psychotherapist. But, as said earlier, the conditions lacked the modern age instances. So, we went through forums and blogs where depressed people shared their problems. From there, we deduced some grossing problems that seemed to be very common and added those conditions accordingly based on regular expressions, matching certain keywords. The below two screenshots shows the conditions which were taken from Eliza and the conditions which we added from the forums.

```
["Why do you need {0}?",
 "Would it really help you to get {0}?",
 "Are you sure you need {0}?"],

[r'Why don\'t you ([^\\?]*)\?\?',
 ["Do you really think I don't {0}?",
 "Perhaps eventually I will {0}.",
 "Do you really want me to {0}?"],

[r'Why can\'t I ([^\\?]*)\?\?',
 ["Do you think you should be able to {0}?",
 "If you could {0}, what would you do?",
 "I don't know -- why can't you {0}?",
 "Have you really tried?"],

[r'I can\'t (.*)',
 ["How do you know you can't {0}?",
 "Perhaps you could {0} if you tried.",
 "What would it take for you to {0}?"],

[r'I am (.*)',
 ["Did you come to me because you are {0}?",
 "How long have you been {0}?",
 "How do you feel about being {0}?"],

[r'I\'m (.*)',
 ["How does being {0} make you feel?",
 "Do you enjoy being {0}?",
 "Why do you tell me you're {0}?",
 "Why do you think you're {0}?"],
```

[Source: Our project program]

```
[r'(.*) children(.*)',
 ["Do you think you are close enough with your children?",
 "Are your children stressed with their personal lives?",
 "Do you remember the last quality time you spent with your children? Can you tell me about it?",
 "Don't you think you should talk to your children freely about the problems you are facing?"],

[r'(.*) friend(.*)',
 ["Tell me more about your friends."
 "When you think of a friend, what or who comes to your mind?",
 "Do you have close friends or did you have one? Then tell me about that",
 ],

[r'(.*) job(.*)',
 ["Tell me more about your job."
 "What problems are you facing at your job?",
 "Do you have any close friend at job? Did you try to speak to him/her about this?",
 "Are you completely satisfied with the job that you are doing? Do you think you deserve better?",
 ],

[r'(.*) wife(.*)',
 ["Tell me more about your wife."
 "What exact problems are you facing with your wife?",
 "Do you think your wife is having any external affairs?",
 "Did you try talking to your wife about this current scenario?",
 "Do you think your wife is hiding something from you? Are you certain? If yes, why do you feel so?"
 ],
```

[Source: Our project program]

Then, we write two functions, one to take in consideration, the reflections list that we imported before. This function will change match input text pattern and change the you to I and vice-versa in accordance with the input tokens. And the second function is for analyzing the input chat. It will analyze the input words and match it with the keywords in the list of regular expressions. If it matches, it will randomly generate any of the conditions attached to that certain regular expression. Like this, a conversation will go on till the user decides to quit. The token for quitting the application is the string “good-bye”. When someone types “good-bye” in any of the input, the program will do a checking based on the requirement of the classifier. The classifier needs at least 30 words as input to perform the supervised learning. So, upon typing “good-bye” the program will count the total number of input words. If the condition satisfies, the program will exit with a message, else it will ask the user to continue chatting a bit more. The input chat gets stored in a string variable and that string is fed to the classifier program. An instance of the conversational flow of the chatbot is shown below.

```
C:\Users\lenovo pc\Anaconda3>python mybot.py
Hi, Please express your answers in sentences and not in one or two words!
We will be happy to help you :)!
And if you want to exit, just type 'good-bye'
Hello. How are you feeling today?
> I am sad
Did you come to me because you are sad?
> Yes
OK, but can you elaborate a bit?
> I think my wife is cheating on me
Do you really think so?
> good-bye

Need to chat a bit more to help us analyze. Thank You!
So. Tell me a bit more about your problem
> My wife is not being loyal to me
I see, your wife is not being loyal to you.
> Yes.
OK, but can you elaborate a bit?
> I think she is cheating on me
Do you really think so?
> good-bye

Analyzing your responses...
```

[Source: Our project program]

The second phase part of the project is the supervised learning. Before writing the python script for classification, we built two corpuses for training the machine algorithms; One corpus comprised of the happy or non-depressed words which we obtained by referring a paper titled HappyDB, published in January 2018^[4]. And for the second corpus, it comprised of all the depressing words, which we built referring a reddit forum^[5]. Based on these words in the corpuses, the training model of the classifier will be built and that model will be used to classify the input text string that will be fed to the rogram. The below screenshots show instances of depressed and non-depressed corpuses respectively.

die I need to die I need to die I need to die I need to die I need to die I need to die I need to die I need to die I don't know how I'd ever say any of this to any person in real life, they'd probably think I'm either crazy or lying, and I know how hard some people have it in life, I don't mean to try and say my life is the worst, I know it can always be worse.. That being said I really need to say some things and I don't really know of a more supportive community on Reddit then r/depression so maybe making this will help, who knows, who cares, it honestly doesn't matter that much. I've had depression for my entire life it seems, I can't remember a time when I haven't felt this way but I know my state is depression because it's gotten worse over the years. I'm not going into great detail because no one wants to read that, not even I do so I'll keep my memories brief. My first memory was my dad firing a gun, I remember the flash, bang, and ringing in my ears. Then my childhood seemed normal. Then my dad molested me. Then my childhood seemed normal but he felt off and I was sad but didn't know why. Then my dad raped my sister and made me watch. Then my childhood seemed normal but I felt differently detached and life seemed less living. Then my dad made me suck my brothers penis. Then my childhood seemed normal but I felt guilty. Then I got beaten by my dad. Then I my childhood seemed normal but by then I was in the darkness, I'm still searching for the light. I wish I hadn't been so young, I wish I had known how to fight back, how to tell, how to scream out for help, how to face evil and scream out to the world that I had seen evil. But I never said anything, I didn't know any better but knew these things were wrong. After these things in the short period of time they occurred I was overwhelmed. I lived my life, I grew up, I went to school and did the normal things people do to be normal. I acted fine around everyone and maybe sometimes I believed I could fit in. But I never opened up, I was never vulnerable to others so I spent my life isolated besides the regular small friendgroup and school. I missed out on a lot of what would constitute as normal experience because of that, wish I hadn't been so isolated both in my shell of human interaction and my literal isolation and coping through addiction with boring

[Source: Our project program]

27677,6227,24h,"I went with grandchildren to butterfly display at Crohn Conservatory",
 ",I went with grandchildren to butterfly display at Crohn Conservatory",
 ",True,1,,affection
 27678,45,24h,I meditated last night.,I meditated last night.,True,1,leisure,leisure
 27679,195,24h,"I made a new recipe for peasant bread, and it came out spectacular!","I made a new recipe for peasant bread, and it
 27680,740,24h,I got gift from my elder brother which was really surprising me,I got gift from my elder brother which was really s
 27681,3,24h,YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED,YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED,True,1,,enjoy_the_moment
 27682,4833,24h,Watching cupcake wars with my three teen children,Watching cupcake wars with my three teen children,True,1,,affect
 27683,7334,24h,I came in 3rd place in my Call of Duty video game.,I came in 3rd place in my Call of Duty video game.,True,1,,leis
 27684,78,24h,I completed my 5 miles run without break. It makes me feel strong.,I completed my 5 miles run without break. It make
 27685,21,24h,went to movies with my friends it was fun ,went to movies with my friends it was fun ,True,1,,bonding
 27686,8,24h,I was shorting Gold and made \$200 from the trade.,I was shorting Gold and made \$200 from the trade.,True,1,,achieveme
 27687,586,24h,"Hearing Songs It can be nearly impossible to go from angry to happy, so youare just looking for the thought that e
 the direction of happiness. It may take a while, but as long as you're headed in a more positive direction youall be doing yourse
 27688,489,24h,My son performed very well for a test preparation.,My son performed very well for a test preparation.,True,1,,affect
 27689,976,24h,I helped my neighbour to fix their car damages. ,I helped my neighbour to fix their car damages. ,True,1,,bonding
 27690,3972,24h,Managed to get the final trophy in a game I was playing. ,Managed to get the final trophy in a game I was playing.
 27691,14,24h,A hot kiss with my girl friend last night made my day,A hot kiss with my girl friend last night made my day,True,1,,
 27692,1230,24h,My new BCAAs came in the mail. Yay! Strawberry Lemonade flavored aminos make my heart happy.,My new BCAAs came in
 27693,9950,24h,Got A in class.,Got A in class.,True,1,,achievement
 27694,4862,24h,My sister called me from abroad this morning after some long years. Such a happy occassion for all family members.
 27695,4891,24h,The cake I made today came out amazing. It tasted amazing as well.,The cake I made today came out amazing. It tast
 27696,737,24h,"There are two types of people in the world: those who choose to be happy, and those who choose to be unhappy. Cont
 27697,498,24h,My grandmother start to walk from the bed after a long time.,My grandmother start to walk from the bed after a long
 27698,2124,24h,i was able to hit a top spin serve in tennis,i was able to hit a top spin serve in tennis,True,1,,achievement
 27699,156,24h,I napped with my husband on the bed this afternoon and it was sweet to cuddle so close to him.,I napped with my hus
 27700,1783,24h,My co-woker started playing a Carley Rae Jepsen song from her phone while ringing out customers.,My co-woker start
 27701,3382,24h,"My son woke me up to a fantastic breakfast of eggs, his special hamburger patty and pancakes. ",
 "My son woke me u

[Source: Our project program]

After the corpus is built, the next thing that we did was text pre-processing. The input text that is fed to the program is pre-processed. The methods used are:

- Tokenization
- Removing punctuations

- Stemming/Lemmatization
- Removing any numbers, special characters etc.

The pre-processed lists of words were then captured in data frames. Then we utilized these data frames to build a unigram and bigram model. The data frames had data labelled as depressed and not depressed. After that, we utilized TFDIF Vectorizer and Count Vectorizer on these words to calculate weights. Using the train_test_split in scikitlearn package of python, we split the data into two parts and then trained our model using the training data. We then predicted our labels using testing data and then we calculated the model accuracy based on predicted labels and compared them with test labels. We also calculated other metrics such as Precision, F measure, and Recall. The data from the chatbot is then fed to the model and then the labels are predicted. Based on the count of the no of labels, we can tell if the person is depressed or not depressed. The screenshots below show how the data was captured in our data frames:

Bigram-Depressed

	Data	Category
0	im year	Depressed
1	year old	Depressed
2	old life	Depressed
3	life feel	Depressed
4	feel like	Depressed
5	like wast	Depressed
6	wast time	Depressed
7	time purpos	Depressed
8	purpos motiv	Depressed
9	motiv ive	Depressed
10	ive diagnos	Depressed
11	diagnos depress	Depressed
12	depress anxieti	Depressed
13	anxieti think	Depressed
14	think might	Depressed
15	might mild	Depressed
16	mild case	Depressed
17	case two	Depressed
18	two dont	Depressed
19	dont know	Depressed
20	know im	Depressed
21	im law	Depressed
22	law graduat	Depressed
23	graduat didnt	Depressed
24	didnt get	Depressed
25	get impress	Depressed
26	impress grade	Depressed
27	grade work	Depressed
28	work retail	Depressed
29	retail legal	Depressed

[Source: Our project program]

Bigram-Not Depressed

	Data	Category
0	hi happi	NotDepressed
1	happi son	NotDepressed
2	son got	NotDepressed
3	got mark	NotDepressed
4	mark examin	NotDepressed
5	examin happi	NotDepressed
6	happi son	NotDepressed
7	son got	NotDepressed
8	got mark	NotDepressed
9	mark examin	NotDepressed
10	examin trueeffect	NotDepressed
11	trueeffect hi	NotDepressed
12	hi went	NotDepressed
13	went gym	NotDepressed
14	gym morn	NotDepressed
15	morn yogai	NotDepressed
16	yogai went	NotDepressed
17	went gym	NotDepressed
18	gym morn	NotDepressed
19	morn yogatrueexercis	NotDepressed
20	yogatrueexercis hwe	NotDepressed
21	hwe seriou	NotDepressed
22	seriou talk	NotDepressed
23	talk friend	NotDepressed
24	friend flaki	NotDepressed
25	flaki late	NotDepressed
26	late understood	NotDepressed
27	understood good	NotDepressed
28	good even	NotDepressed
29	even hang	NotDepressed

[Source: Our project program]

Unigram-Depressed

	Data	Category
0	im	Depressed
1	year	Depressed
2	old	Depressed
3	life	Depressed
4	feel	Depressed
5	like	Depressed
6	wast	Depressed
7	time	Depressed
8	purpos	Depressed
9	motiv	Depressed
10	ive	Depressed
11	diagnos	Depressed
12	depress	Depressed
13	anxieti	Depressed
14	think	Depressed
15	might	Depressed
16	mild	Depressed
17	case	Depressed
18	two	Depressed
19	dont	Depressed
20	know	Depressed
21	im	Depressed
22	law	Depressed
23	graduat	Depressed
24	didnt	Depressed
25	get	Depressed
26	impress	Depressed
27	grade	Depressed

[Source: Our project program]

Unigram-Not Depressed

	Data	Category
0	hi	NotDepressed
1	happi	NotDepressed
2	son	NotDepressed
3	got	NotDepressed
4	mark	NotDepressed
5	examin	NotDepressed
6	happi	NotDepressed
7	son	NotDepressed
8	got	NotDepressed
9	mark	NotDepressed
10	examin	NotDepressed
11	trueaffect	NotDepressed
12	hi	NotDepressed
13	went	NotDepressed
14	gym	NotDepressed
15	morn	NotDepressed
16	yogai	NotDepressed
17	went	NotDepressed
18	gym	NotDepressed
19	morn	NotDepressed
20	yogatrueexercis	NotDepressed
21	hwe	NotDepressed
22	seriou	NotDepressed
23	talk	NotDepressed
24	friend	NotDepressed
25	flaki	NotDepressed
26	late	NotDepressed

[Source: Our project program]

The below screenshot shows an instance of our classification program,

```
def main():
    input3=0
    input = int(raw_input("Select The following 1.Unigram 2.bigram"))
    input1 = int(raw_input("Select The following 1.TfidfVectorizer 2.CountVectorizer"))
    input2 = int(raw_input("Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Rando
Depressedvalues = readFile('/Users/rohitgs/Desktop/Depressed.txt')
NotDepressedvalues = readFile('/Users/rohitgs/Desktop/Not-Depressed.txt')
Depressedvalues = preprocess(Depressedvalues)
bigram_list=[]
bigram = True
bigram_list_Depressed = []
bigram_list_NotDepressed = []
NotDepressedvalues = preprocess(NotDepressedvalues)
Depressed = []
NotDepressed = []
Depressed.append('Depressed')
NotDepressed.append('NotDepressed')
if input == 1:
    Depressed =Depressed * len(Depressedvalues)
    NotDepressed = NotDepressed *len(NotDepressedvalues)
    df2 = pd.DataFrame({'Category':Depressed})
    df1 = pd.DataFrame({'Data':Depressedvalues})
    Depressed_dataframe = df1.join(df2 )
    print Depressed_dataframe
    df4 = pd.DataFrame({'Category':NotDepressed})
    df3 = pd.DataFrame({'Data':NotDepressedvalues})
    NonDepressed_Dataframe = df3.join(df4)
    print NonDepressed_Dataframe
    Total_Dataframe = Depressed_dataframe.append(NonDepressed_Dataframe)
    print Total_Dataframe
    #print Total_Dataframe
    if (input1 == 1):
        Vectorizer = TfidfVectorizer(ngram_range=(1, 1), min_df=2, stop_words='english')
```

[Source: Our project program]

We used several machine learning algorithms to see which suited our model the best. To brief on the machine learning algorithms utilized in this program, we used:

- Multinomial Naïve Bayes – Naïve Bayes classification is probabilistic classification based on Bayes' Theorem and assuming strong independence between the features^[6]. Multinomial event model is the model which is suitable for classification with discrete features (e.g. word count for text classification). The probability equation of Multinomial Naïve Bayes is given by:

$$p(\mathbf{x} \mid C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{ki}^{x_i}$$

[Source: https://en.wikipedia.org/wiki/Naive_Bayes_classifier]

- Gaussian Naïve Bayes – Gaussian Naïve Bayes is used when dealing with continuous data. The assumption is that the continuous values of each class follow a Gaussian distribution^[6]. The probability equation of Gaussian Naïve Bayes is given by:

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

[Source: https://en.wikipedia.org/wiki/Naive_Bayes_classifier]

- SVM for Linear Kernel – SVM or Support Vector Machines are supervised learning models that analyze the data used for classification and regression analysis^[7]. The hyperplane equation of linear SVM model is represented as:

$$\vec{w} \cdot \vec{x} - b = 0,$$

[Source: https://en.wikipedia.org/wiki/Support_vector_machine]

where w is the normal vector to the hyperplane.

- Random Forest Classifier – Random Forest Classifier is an ensemble learning method for classification that fits decision tree classifiers on various sub-samples of the dataset^[8]. It uses averaging to improve the accuracy. This classifier fixes the over-fitting problem of the Decision Trees. After training the model, predictions for unseen data is done by averaging the predictions from the individual regression trees which is represented as:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x')$$

[Source: https://en.wikipedia.org/wiki/Random_forest]

- MLP Classifier – MLP or Multilayer Perceptron is a class of feedforward artificial neural network^[9]. It uses backpropagation for training technique to implement supervised learning. It has multiple layers and it has the property of non-linear activation^[9]. Learning takes place by changing connection weights after each piece of data is processed which is represented as:

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n)$$

[Source: https://en.wikipedia.org/wiki/Multilayer_perceptron]

IV. EXPERIMENT DESIGN

After applying all the classification algorithms, it is very much necessary to test our model so that can measure the performance and use it to classify unseen data. So, for calculating performance, we took five parameters into consideration, Accuracy, Precision, Recall, F1-Score and Accuracy:

- Accuracy – Accuracy score is the percentage calculation of the correct predictions. The mathematical formula is given by:

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i)$$

[Source: http://scikit-learn.org/stable/modules/model_evaluation.html]

- Precision – Precision is defined in terms of retrieved documents and relevant documents and the formula for the same is given by^[10]:

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

[Source: https://en.wikipedia.org/wiki/Precision_and_recall]

- Recall – Similarly like Precision, Recall is also defined in terms of retrieved and relevant documents, and the formula is given by:

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

[Source: https://en.wikipedia.org/wiki/Precision_and_recall]

- F1 Score – F1 score is also a measure for test's accuracy^[11]. It takes into consideration both Precision and Recall to compute the harmonic average of the two. The formula for F1 score is given by:

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

[Source: https://en.wikipedia.org/wiki/F1_score]

- Support – Support is nothing but the measure of the number of samples of true response present in the particular class. It can be obtained from metrics library of sklearn in python.

So, we calculated all these parameters for all the classification algorithms we implemented. Below screenshots show the performance metrics of each of the algorithms:

- Multinomial Naïve Bayes :-

Multinomial-Bigram-CountVectorizer

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier2
0.895638629283
precision    recall  f1-score   support

   Depressed    0.89    0.92    0.91     783
  NotDepressed    0.90    0.86    0.88     501

 avg / total    0.90    0.90    0.90    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

Multinomial-Bigram-Tfidf

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier2
0.901869158879
precision    recall  f1-score   support

   Depressed    0.88    0.94    0.91     663
  NotDepressed    0.93    0.86    0.89     621

 avg / total    0.90    0.90    0.90    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

Multinomial-CountVectorizer-Unigram

```
Rahuls-MacBook-Air:Desktop rohitis$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier2
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.750778016199

      precision    recall  f1-score   support

   Depressed      0.71      0.90      0.79      678
  NotDepressed      0.84      0.58      0.69      686

 avg / total      0.77      0.75      0.74     1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohitis$
```

Multinomial-Tfidf-Unigram

```
Rahuls-MacBook-Air:Desktop rohitis$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier2
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.756230529595

      precision    recall  f1-score   support

   Depressed      0.72      0.90      0.80      702
  NotDepressed      0.83      0.59      0.69      502

 avg / total      0.77      0.76      0.75     1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohitis$
```

- Gaussian Naïve Bayes :-

Gaussian-Bigram-Tfidf

```
Rahuls-MacBook-Air:Desktop rohitis$ python Script1.py
Select The following 1.Unigram 2.bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier4
0.914330218869

      precision    recall  f1-score   support

   Depressed      0.87      0.99      0.93      708
  NotDepressed      0.99      0.82      0.90      576

 avg / total      0.92      0.91      0.91     1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohitis$
```

Gaussian-CountVectorizer-Unigram

```
Rahuls-MacBook-Air:Desktop rohitis$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier4
0.718847352025

      precision    recall  f1-score   support

   Depressed      0.66      0.96      0.78      674
  NotDepressed      0.91      0.45      0.60      610

 avg / total      0.78      0.72      0.70     1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohitis$
```

Gaussian-CountVectorizer-Bigram

```
Rahuls-MacBook-Air:Desktop rohitis$ python Script1.py
Select The following 1.Unigram 2.bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier4
0.921339563063

      precision    recall  f1-score   support

   Depressed      0.88      0.99      0.93      687
  NotDepressed      0.98      0.85      0.91      597

 avg / total      0.93      0.92      0.92     1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohitis$
```

Gaussian-Tfidf-Unigram

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier4
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.70815576324
precision    recall  f1-score   support

   Depressed      0.63     0.95     0.76     641
  NotDepressed     0.91     0.45     0.60     643

 avg / total      0.77     0.70     0.68    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

- SVM Classifier :-

SVM-Bigram-Tfidf

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier1
0.90106919079
precision    recall  f1-score   support

   Depressed      0.89     0.93     0.91     664
  NotDepressed     0.92     0.87     0.90     620

 avg / total      0.90     0.90     0.90    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

SVM-Countvectorizer-Bigram

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier1
0.894059013084
precision    recall  f1-score   support

   Depressed      0.87     0.93     0.90     657
  NotDepressed     0.92     0.86     0.89     627

 avg / total      0.90     0.89     0.89    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

SVM-CountVectorizer-Unigram

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier1
0.759345794393
precision    recall  f1-score   support

   Depressed      0.72     0.88     0.79     672
  NotDepressed     0.83     0.62     0.71     612

 avg / total      0.77     0.76     0.75    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

SVM-Tfidf-Unigram

```
Rahuls-MacBook-Air:Desktop rohittg$ python Script1.py
Select The following 1.Unigram 2.bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPCClassifier1
0.761682242991
precision    recall  f1-score   support

   Depressed      0.73     0.90     0.80     704
  NotDepressed     0.83     0.60     0.69     580

 avg / total      0.77     0.76     0.75    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittg$
```

- Random Forest Classifier :-

RandomForest-CountVectorizer-Unigram

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram]
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier3
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.759345794393
precision    recall  f1-score   support

   Depressed    0.74    0.86    0.79     684
  NotDepressed    0.88    0.65    0.72     698

 avg / total    0.76    0.76    0.76    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

RandomForest-CountVectorizer-Bigram

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier3
0.854361370717
precision    recall  f1-score   support

   Depressed    0.84    0.87    0.86     645
  NotDepressed    0.87    0.84    0.85     639

 avg / total    0.85    0.85    0.85    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

RandomForst-Bigram-Tfdif

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier3
0.89953271026
precision    recall  f1-score   support

   Depressed    0.88    0.94    0.91     671
  NotDepressed    0.93    0.86    0.89     613

 avg / total    0.90    0.90    0.90    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

RandomForest-Tfdif-Unigram

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier3
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a l
ot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.750778816199
precision    recall  f1-score   support

   Depressed    0.73    0.85    0.78     676
  NotDepressed    0.79    0.64    0.71     698

 avg / total    0.76    0.75    0.75    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```


- MLP Classifier :-

MLP-Bigram-Tfidf

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier5
0.915109034268
precision    recall  f1-score   support

   Depressed    0.91    0.94    0.92     690
  NotDepressed    0.93    0.89    0.91     594

 avg / total    0.92    0.92    0.91    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

MLP-Bigram-CountVectorizer

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram2
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier5
0.909657328872
precision    recall  f1-score   support

   Depressed    0.91    0.92    0.91     658
  NotDepressed    0.91    0.98    0.91     626

 avg / total    0.91    0.91    0.91    1284

Input String
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

MLP-CountVectorizer-Unigram

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer2
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier5
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.743769470405
precision    recall  f1-score   support

   Depressed    0.71    0.86    0.77     661
  NotDepressed    0.80    0.62    0.70     623

 avg / total    0.75    0.74    0.74    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

MLP-Tfidf-Unigram

```
Rahuls-MacBook-Air:Desktop rohittgs$ python Script1.py
Select The following 1.Unigram 2.Bigram1
Select The following 1.TfidfVectorizer 2.CountVectorizer1
Select The Machine Learning algorithm 1.SVM 2.MultinomialNB 3.Random Forest 4.Gaussian naive bayes 5.MLPClassifier5
Given Input
Hi i am not feeling very good today. It makes me feel very bad inside. I feel i should die I used to be a rich person but my wife stole all my money and left for another man We used to love each other a lot. But i started doubting him with another man One day at night i heard her speaking with some one else on the call. I have no friends I dont know. Yes I quit

Accuracy
0.741439021807
precision    recall  f1-score   support

   Depressed    0.72    0.83    0.77     669
  NotDepressed    0.78    0.64    0.70     615

 avg / total    0.75    0.74    0.74    1284

The person is depressed
Rahuls-MacBook-Air:Desktop rohittgs$
```

V. CONCLUSION

Looking onto the present scenario and the current statistics of young people getting depressed, we aimed to develop a tool which can help reduce this severe problem. In doing so, we went through multiple research papers to gain a knowledge on the works that has already been done on this area. Based on the knowledge we acquired we started building the application, keeping two such papers as the backbone of our application, the details of which has been given in the “Related Work” header of the report.

We chose Python as our coding language platform. We built a chatbot, keeping in mind, the existing chatbot Eliza. As Eliza was built with the help of a psychotherapist, we took the therapy conditions from it and added some more to it, where it lacked. Then we stored the conversation and passed the text to our classification program which classified the text based on the existing corpuses. So, as we can see the screenshots in the “Experiment Design” header, SVM performed the best for our unigram model, while for the bigram model for Gaussian Naïve Bayes performed the best with accuracy score of 0.92 as compared to other classifiers. So, we analyzed the person’s text and classified him as depressed or not-depressed.

Our application aims to solve a grossing real-world problem of 21st century. We hope that by our work, we can help people and decrease the suicide rates which has been rising exponentially every year.

REFERENCES

- [1] Anon. 2017. July 2017: Youth Suicide. (August 2017). Retrieved April 14, 2018 from <https://www.relationships.org.au/what-we-do/research/online-survey/july-2017-youth-suicide>
- [2] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 1 (January 1966), 36-45. DOI=<http://dx.doi.org/10.1145/365153.365168>
- [3] Nadeem and Moin. 2016. Identifying Depression on Twitter. (July 2016). Retrieved April 14, 2018 from <https://arxiv.org/abs/1607.07384>
- [4] Asai et al. 2018. HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments. (January 2018). Retrieved April 14, 2018 from <https://arxiv.org/abs/1801.07746>
- [5] Anon. /r/depression, because nobody should be alone in a dark place. Retrieved April 14, 2018 from <https://www.reddit.com/r/depression/>
- [6] Anon. 2018. Naive Bayes classifier. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/Naive_Bayes_classifier
- [7] Anon. 2018. Support vector machine. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/Support_vector_machine
- [8] Anon. 2018. Random forest. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/Random_forest
- [9] Anon. 2018. Multilayer perceptron. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/Multilayer_perceptron
- [10] Anon. 2018. Precision and recall. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/Precision_and_recall
- [11] Anon. 2018. F1 score. (April 2018). Retrieved April 14, 2018 from https://en.wikipedia.org/wiki/F1_score

APPENDICES

Eliza – Psychotherapist chatbot developed in 1966 by Joseph Weizenbaum

MLP – Multi Layer Perceptron

SVM – Support Vector Model

Tfidf – Term Frequency and Inverse Document Frequency

Project Code - <https://svn.cs.dal.ca/nlp-course/P-02/Project-Source-Code/>
