

ChatBot which Identifies Depression [P-02]:

Names of the Members in the Group:

Rohit Gollarahalli

Nibir Mukherjee (B00768437)

Problem Statement:

Depression is a very common mood disorder, identifying such illness is important as many cases in today's world go undiagnosed. Texting has been a modern form of communication in today's world. With current shortage of available psychiatric professionals, we see a steep rise in anxiety and depression among today's youngsters^[5]. To deal with this, we intend to build a chatbot that aims to identify whether a person is suffering from depression using machine learning algorithms.

List of Possible approaches:

The chatbot which we will built will engage in conversation with the user. Based on the words typed and inputs given, the chatbot can identify and classify the person as "depressed" or "not depressed". To deal with this we first design our chatbot. Chat Bot will be designed based on pattern matching. The conversational agent was first implemented in 1966 by Joseph Weizenbaum^[1].

On the part of identifying depression, we will see it particularly as a classifier problem^[2]. We will be designing a corpus from various forums. This will have a combined list of both depressing words and non-depressing/cheerful words^[8].

We will be training our model using this corpus. The Input text that is entered in the Chatbot will be collected and fed to the model to identify the category it comes under.

We intend to build corpus with our own data which will have set of depressing words.

To Improve our predictions, we will be pre-processing the data:

1. Tokenizing
2. Removing the stop words: English
3. Stemming/Lemmatization
4. Removing any punctuation, numerical, special characters and lowering case

We intend to convert data into sparse matrix, feed it to a pipeline and then fit into a model.

We would be using the following algorithms's

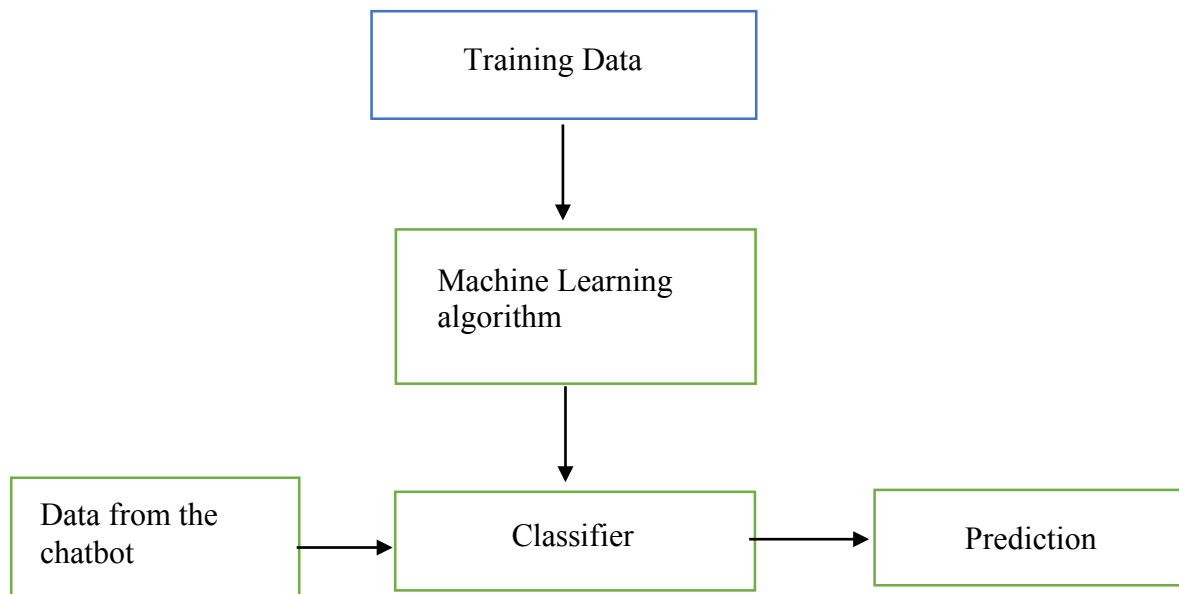
1. Multinomial Naive Bayes
2. SVM classifier [for various kernel: linear, poly, rbf, sigmoid]
3. Random Forest

We have plans of implementing other classifiers too and calculate the accuracy. (For eg, Logistic regression, Multilayer perceptron and Decision tree)

We plan to use the following libraries:

1. sklearn
2. NLTK

Below is the flow diagram:



Project plan for the rest of the term

<i>Date</i>	<i>Action Items/Tasks</i>
March 21st:	Building the Corpus
March 27 th	Completing the Chatbot
April 2 nd	Applying Machine learning algorithms and calculating their accuracy.
April 5 th	Completing project Report and Slides
April 6 th	Presentation

List of References:

- [1] Weizenbaum, Joseph "ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine" in: Communications of the ACM; Volume 9 , Issue 1 (January 1966): p 36-45.
- [2]Nadeem and Moin. 2016. Identifying Depression on Twitter. (July 2016). Retrieved March 19, 2018 from <https://arxiv.org/abs/1607.07384>
- [3] Anon. Dr. Sebastian Raschka. Retrieved March 19, 2018 from <https://sebastianraschka.com/>
- [4] Anon. Working With Text Data¶. Retrieved March 19, 2018 from http://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html
- [5]Anon. 2018. Chatbot: The Therapist in Your Pocket. (January 2018). Retrieved March 19, 2018 from <https://iq.intel.com/chatbots-the-therapist-in-your-pocket/>
- [6]Will Knight. 2017. A new chatbot will help with depression-and it's backed by one of the biggest names in AI. (October 2017). Retrieved March 19, 2018 from <https://www.technologyreview.com/s/609142/andrew-ng-has-a-chatbot-that-can-help-with-depression/>
- [7]Anon. /r/depression, because nobody should be alone in a dark place. Retrieved March 19, 2018 from <https://www.reddit.com/r/depression/>
- [8] Vik Paruchuri. 2018. Naive bayes: Predicting movie review sentiment. (February 2018). Retrieved March 19, 2018 from <https://www.dataquest.io/blog/naive-bayes-tutorial/>
- [9] Asai et al. 2018. HappyDB: A Corpus of 100,000 Crowdsourced Happy Moments. (January 2018). Retrieved March 19, 2018 from <https://arxiv.org/abs/1801.07746>