International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206527

Investigating differential linguistic patterns exhibited by Major Depressive Disorder (MDD) Patients and building a Long Short Term Memory Network + Convolutional Neural Network Model, Logistic Regression model, and a Multinomial Naive Bayes Classifier Algorithm to develop Spero, a hybrid app based Early-MDD diagnosis system

Shivam Garg, Ashley Raigosa, Rimsha Aiman
SVKM International School, Mumbai, Maharashtra, India

Article Info
Volume 6, Issue 5
Page Number: 114-127
Publication Issue :
September-October-2020

Article History
Accepted : 20 Sep 2020
Published : 30 Sep 2020

ABSTRACT

Major Depressive Disorder (MDD), otherwise known as Depression, is the leading psychiatric disorder globally in terms of the number of individuals it affects. Despite this there is no effective and reliable early diagnostics system for MDD. Hence, through this study, we aimed to fill this void by not only investigating linguistic differences in posts made on social media by people exhibiting and people not exhibiting symptoms of MDD but also by developing various machine learning architectures to build an accessible, sensitive, and accurate MDD early diagnostics system. Through the differential linguistic analysis we conducted on the dataset we manually scraped and filtered, we clearly demonstrated that there indeed were certain linguistic and topical features that were different amongst depressed and healthy patients. Furthermore, we also successfully built three different ML Algorithms in which our Long Short Term Memory Network (LSTM) + Convolutional Neural Network (CNN) Model attained an accuracy of 95.00%, our Multinomial Naive Bayes Classifier Algorithm attained an accuracy of 92%, and our Logistic Regression Model achieved an accuracy of 87.627%. Ultimately, given the LSTM + CNN Model's high accuracy, weighted precision (0.95), recall (0.95), and f-1 score (0.95), we decided to integrate it into an app built on Swift UI to develop Spero, a first of its kind early diagnostics system for MDD.

Keywords : Linguistic Analysis, Language Modeling, Sentimental Analysis, Major Depressive Disorders, Early Diagnostic Systems, Long Short Term Memory Network, Convolutional Neural Network, Logistic Regression Model, Multinomial Naive Bayes Algorithm, Depression

I. INTRODUCTION

Major Depressive Disorder (MDD), otherwise known as Depression, is the leading psychiatric disorder globally in terms of the number of individuals it affects. According to the World Health Organisation (WHO), over 264 million people of all ages suffer from depression and it is the leading cause of

disability worldwide [1]. Depression is a significant medical condition that affects several areas of a person's life and elicits short-lived and intense feelings of emotions, primarily sad, for extended durations of time [2]. MDD often leads to suicide and close to 800,000 people die due to it every year, making it the second leading cause of death amongst the 15-29 years old age bracket.

Various Institutions like the Institution of Medicine Committee on the Prevention of Mental Disorders have identified depression as a preventable disorder [3], and several studies have also proven that its treatment can mitigate its ill effects [4-5] and despite known effective treatments, between 76% to 85% of people residing in low and middle-income countries receive no treatment whatsoever for their disorder [6]. Barriers to effective care for patients suffering from such disorders include lack of trained health-care providers, lack of resources, and the general social stigma associated with mental health disorders. More importantly though, the lackof effective care is also due to inaccurate assessments and the lack of early diagnosis systems. Therefore, it is quintessential to develop early diagnostics assays and techniques to potentially intervene and halt any further escalation. However, such provisions remain scarce to date. Although there exist validated laboratory tests to diagnose depression, such as The Hamilton scale for rating depression [7], Geriatric depression scale, Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Hospital anxiety and depression scale, and Patient Health Questionnaire [8,9], they are rendered ineffective as most diagnoses are based on self-reports through questionnaires.

Hence, in this study, we decided to investigate differences in linguistic patterns in posts and messages on social media between people exhibiting symptoms of MDD (experimental group) and people not exhibiting symptoms of MDD to ultimately use various machine learning architectures to develop an accessible, reliable, and accurate MDD early diagnostics system.

II. Literature Review

The association between language and clinical disorders has been studied for years [10]. Studies have been conducted to predict clinical disorders using user activity on various social media sites like Twitter, Reddit, Instagram, etc. For instance, Prieto et al [11] used twitter to automatically detect the incidence of a set of health conditions. Even Aladağ et al [12] studied different posts for their linguistic patterns to prevent people from committing suicides. In fact, Leis et al [13] even noticed differences in general characteristics of language and user activity in depressed users of social media, specifically Twitter, when compared to those not suffering from this disorder. Hence, social media holds great potential for researchers for analyzing a whole host of medical conditions as by studying huge amounts of text in posts and comments, researchers can link language use with social behavior and personality [14].

Certain indicators (such as first-person pronouns) of Self-Focused Attention (SFA), a cognitive bias that is closely related to anxiety and depression, were also found to be positively related with varying signs of co-morbid depression and anxiety. Furthermore, it was also noted that the use of first-person pronouns is greater during negative memory recall than during positive memory recall. Another study [16] led by Seabrook et al conducted a longitudinal analysis of Facebook and Twitter Status Updates of 29 Facebook users and 49 Twitter users and found that the use of negative words was correlated to the intensity of depression. Apart from this, it has also been noted that depressed individuals are found to post content more frequently [17] and that changes in the severity of depression may be indicated by increases in posting behavior on Social media. Hence, all previous

studies point to how several features of the social media activity, such as the number, relative frequency, and temporal distribution of tweets can be used for the detection and monitoring of mental disorders, such as depression [18].

While all studies conducted previously help in establishing a link between language use and the psychological state of mind of a person, no study to date has been conducted to predict depression based on the linguistic patterns that have been identified. Moreover, there is a lack of an easily accessible, cheap, and effective early diagnostics system for depression and no research has yet attempted to develop such a diagnostics system.

Therefore, in this study, we aimed to:

1) Conduct Differential Linguistic Analysis of Depressed and Non-Depressed Individuals 2) Build three different Machine Learning (ML) models for early diagnosis of MDD using linguistic indicators.
3) Develop a novel hybrid app-based system to carry out early diagnosis of MDD in the real world.

III. Methods

To conduct linguistic analysis and train the ML Models we first scraped data from Reddit using PRAW (Python Reddit Api Wrapper) and then mined data from Twitter using Twint. Additionally, we also acquired our data from the Sentiment140 dataset from Kaggle [19]. The data was divided into two groups: experimental group which corresponded to tweets/reddit posts from depressed individuals and control group which corresponded to tweets/reddit posts from healthy patients. This labelled data was manually screened and was used to train the ML models we built. Subsequently, the most accurate ML Model was then exported and integrated into an app on Swift.

Several libraries were used in this project for different purposes. Namely, pandas, tensorflow, sklearn, numpy, keras, matplotlib, nltk, and ftfy were used.

Data Collection

Data was collected from two different Social Media sites, namely Twitter and Reddit, to not only acquire a large dataset but to also acquire a diverse dataset from different types of users. Additionally, having data from two different sites also allowed us to acquire posts of varying lengths as tweets tend to be shorter than reddit posts.

First, the data was scraped from reddit using PRAW. Posts that were scraped from Subreddits like r/depression and r/sucidewatch were labelled 1, indicative of posts made by depressed individuals, and posts from subreddits like r/happy, r/CasualConversation, r/CongratsLikeImFive, r/NeutralPolitics, r/applyingtocollege, and r/unpopularopinion were labelled as 0, indicative of posts made by healthy individuals. It was ensured that the control group had a good distribution of posts from all age brackets, genders, and ethnicities and also covered a wide range of topics to ensure model accuracy. A total of 3906 posts for the experimental group (those indicative of depressed individuals) and 7307 posts for the control group (those indicative of healthy individuals) were scraped from Reddit alone. Subsequently, to ensure the quality of the dataset, each post was manually screened and was checked against the DSM-5 Criteria (Diagnostic and Statistical Manual of Mental Disorders), which are a set of criteria for assessment and diagnosis of mental disorders.

The specific DSM-5 criteria for major depressive disorder are outlined below. [5]. At least 5 of the following symptoms have to be present during the same 2-week period (and at least 1 of the symptoms

must be diminished interest/pleasure or depressed mood) :

1) Depressed mood: For children and adolescents, this can also be an irritable mood 2) Diminished interest or loss of pleasure in almost all activities (anhedonia)
3) Significant weight change or appetite disturbance: For children, this can be failure to achieve expected weight gain
4) Sleep disturbance (insomnia or hypersomnia)
5) Psychomotor agitation or retardation
6) Fatigue or loss of energy
7) Feelings of worthlessness
8) Diminished ability to think or concentrate; indecisiveness
9) Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or specific plan for committing suicide

If posts which were tentatively put in the experimental group, failed against the DSM-5 Criteria, then they were deleted from the dataset. Similarly, a post that met the DSM-5 Criteria in the control group was also deleted. After this manual screening, the filtered Reddit dataset contained a total of 3181 posts for the experimental group and 6762 posts for the control group. Some examples of posts in each group have been given in Appendix B.

Data from twitter was extracted using a different method. While tweets for the experimental group were scraped from Twitter using TWINT and keywords like "depression", the control group wasn't scraped from Twitter, and was instead taken from the Sentiment140 Dataset. After this process we had a total of 12,000 tweets in the control group and 2378 tweets in the experimental group. In both the cases, the dataset for the experimental group was intentionally kept larger to ensure that a wide variety of topics are covered and individuals from different backgrounds are considered.

After the extraction from both Reddit and Twitter, a combined dataset was produced which consisted of 5559 posts/tweets from the experimental group and 18,762 posts/tweets from the control group. We later also subdivided our experimental group into 3 subgroups :

I. Tweets Only
II. Tweets and Reddit Posts both
III. Only Reddit Posts

The control group was kept the same across all these subgroups of the experimental group. The reason behind this subdivision is discussed in later sections.

Linguistic Features

Differential linguistic analysis was conducted to determine the differences in the language used and the topics discussed by depressed and healthy individuals respectively. Specific linguistic features that were looked for were differences in use of first person singular pronouns, i.e "I", and differences in discussion of topics. The entire list of 58 features that were selected can be found in Appendix A. The frequency of discussion of topics like sucide, drugs, body dysmorphia, failing, and life were also looked at and we determined the number of posts from the experimental group and the control group that contained the specific feature using Microsoft Excel. To adjust for the different number of posts in the experimental and control group percentages were calculated. The equation used to calculate the percentage of posts containing a specific linguistic feature for the experimental group is shown below. A similar equation was used for calculating the percentage for the control group.

$$\frac{Number\ of\ posts\ in\ experimental\ group\ contianing\ a\ specific\ linguistic\ feature}{Total\ Number\ of\ posts\ in\ the\ experimental\ group} * 100\%$$

Then, after this was evaluated for both groups for each linguistic feature, a relative presence value was

calculated for each linguistic feature using the formula shown below.

$$Relative\ Presence = \frac{\%\ Posts\ of\ the\ experimental\ group\ containing\ the\ linguistic\ feature}{\%\ Posts\ of\ the\ control\ group\ containing\ the\ linguuistic\ feature}$$

The results obtained and the Relative Presence values were then analysed to see whether certain linguistic features are more prevalent than others in posts made by depressed individuals.

Models

Data Preprocessing

The basic steps of data preprocessing have been outlined below in Figure 1. Initially, the scraped data was cleaned to remove links, images, contractions (do not for don't), hashtags, stopwords, punctuation, emojis, @someone, etc. Then tokenizing, lemmatization, and stemming was carried out. The tokenizer then created an array of every unique word and was applied to the entire dataset. After this an embedding matrix was used to reduce the dimensionality of the data and ultimately, labels were assigned to the depressive posts (experimental group) and non depressive posts (control group), and the data was split into test (60%), validation (20%), and train data (20%), after combining the two groups and randomly shuffling them.

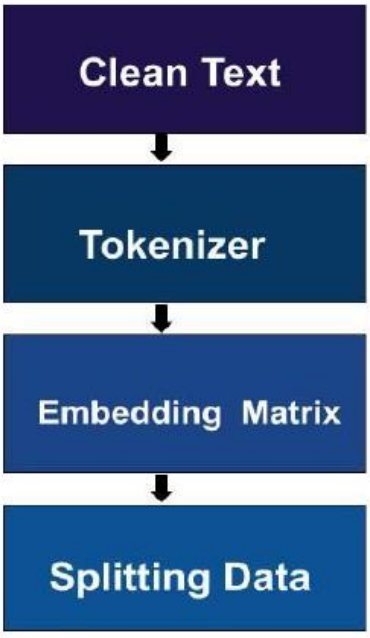


Figure 1 : Steps in Preprocessing Data

Long Short Term Memory Network (LSTM) and Convolutional Neural Network (CNN)

A LSTM + CNN Model was made using Keras to classify a certain post as being indicative of an individual being depressed or not. This was chosen primarily because it allowed for multiclass text classification.

The model architecture has an input layer which receives the tokenized input produced during the preprocessing phase. It is then fed into an embedding vector layer where it gets converted into an embedding vector. This is then run through the Convolutional Neural Network which is adapted to learn the spatial structure of the data from the embedding layer, which is then passed onto an LSTM Layer. The LSTM layer is used for further language modeling and sentimental analysis and its output is then processed by an OPD (Output Dense) Layer which uses the sigmoid function to give the final output. For training the model, 5 epochs were used and an early stopping argument was also employed to prevent overfitting. In summary, the model takes an input text and outputs 1 or 0 signalling whether the post is indicative of depression or not.

Logistic Regression Model

Additional machine learning models were also built to test which model architecture gives the highest accuracy. Hence, a binary classification algorithm, using Logistic Regression was also created as the dependent variable was dichotomous [21]. The logistic regression model was tested with the third subgroup of the experimental group and the entire control group (the reason for this has been described in the results), and the same number of epochs as the LSTM + CNN model allowing this model to also act as a baseline for the previous model.

Multinomial Naive Bayes Classifier Model:

Prior to the creation of the multilayer deep learning model with LSTM and a CNN, we utilized a Multinomial Naive Bayes algorithm to process the data.

The database lacked multiple parameters and allowed for low dimensionality during processing. Thus, the Multinomial Naive Bayes algorithm, which works well with datasets with fewer dimensions, was able to effectively produce accurate results. To implement this algorithm, a count vectorizer was used to tokenize each string and create a vocabulary set for analyzing the data. Next, these results were fitted to the data and transformed to fit the parameters of the model. The train portion utilised 80% of the data while the testing portion utilized 20%. The word vector counts were then passed to a multinomial naive bayes function and each quality of the data was evaluated completely independently with equal weights. To clarify, features like the length of the input text were not incorporated into the results because they did not influence the likelihood of depression.

As a result, the model utilized the training data to evaluate the dataset with limited bias and used binary classification to determine whether the text exhibited signs of depression or not.

App Development

The Spero App makes requests to Twitter and retrieves tweets so that it can be analysed by the inbuilt LSTM and CNN model. To be able to retrieve tweets and the user's twitter information, we used OAuth 1.0. Within the app, OAuth 1.0 creates a url scheme and makes a request to Safari Services to display the login screen for users to enter their account information. Next, an OAuth token is generated which is used to fully log in the user. After this, in order to query for the user's information, an access token is created and an OAuth signature is used to make requests to the Twitter API. In order to retrieve tweets from the user's timeline, including both tweets they've retweeted and posted, TwitterKit is initialized with the authentication credentials from OAuth 1.0 after which it is able to pull these tweets. Then after retrieving the tweets from the user's profile, the tweets are tokenized and embedded through a BOW (Bag of Words) model, as .mlmodel files in Swift can only pass MLMultiArrays, or numbers, and not strings. The BOW model then creates a vector, based on keywords, that is then passed through the LSTM + CNN model. Finally, the output from the LSTM + CNN Model returns the likelihood of the tweet being depressed and only the tweets that are identified as depressed are displayed on the main screen for the user to see. Ultimately, the app is able to make a request for the Twitter user, analyze tweets on their profile, and display the results in a user friendly way.

Implementing this was extremely difficult since TwitterKit and many API's for Twitter have stopped being developed recently, which resulted in them not being compatible with SwiftUI, the most recent version of Swift for iOS development. Hence, a file was created to bridge the functionality of UIKit and a view representable was used to utilize the functions from TwitterKit. This solution solved this problem and after this, Spero worked smoothly and prove to be extremely accurate.

IV. Results and Discussion

Linguistic Features

The conducted analysis, on the third subgroup of the experimental group (reason behind the selection discussed in the next section), revealed some key insights regarding the differences in topics discussed and the language used by depressed and non

depressed individuals. The final results have been shown graphically in Figure 2, and shown in a tabulated manner in Appendix A. "I" had a relative presence of 2.24, clearly indicating that the percentage of depressed posts that contained "I" was two times the percentage of non-depressed posts that contained "I". Hence, this clearly matched what had been found in previous studies as discussed in the Literature Review. Other prominent features were "suicide", which was used 37 times more in depressed posts than in non depressed posts, and "suffer", which was used 20 times more in depressed posts than in non depressed posts. Words like "trapped", too, were used 15 times more, and words like "anxiety" and "useless" were used 12 times more in depressed posts. There was also a greater discussion on bodily features, as words like "ugly", "body", and "fat",which belong to its lexical field, were used 10, 5, and 4 times more respectively in depressed posts than in non-depressed. As expected by the DSM-5 criteria that depressed individuals are more likely to be fatigued and experience appetite disturbance, their language too in their posts reflects this as words like "energy" and "appetite" have been used 9 times more and words like "fatigue" have been used 8 times more in depressed posts. Additionally, as expected, there is also a greater discussion on topics of drugs as words like "smoking", "drugs", "adderall", "alcohol" have been used 2, 5, 3, 5 times more respectively in posts made by individuals with MDD. There is also a greater discussion of life and a general expression of sorrow in posts part of the experimental group, as words like "cry", "die", "struggle", "insomnia", "cares" (in negative context, for instance "nobody cares about me"), "suck", "life", "personality", "bad", "hate", and "like" (in negative context, for instance "nobody likes me"), are used 7, 7, 7, 7, 55, 3, 6, 7, 3, 4, and 118 respectively times more in the experimental group than in the control group. In general, there was also a greater use of negative words like "don't", "can't", and "nothing", as they were used 4, 5, and 7 times more in the depressed posts group than in the nondepressed posts group. Hence, clearly indicating that their social media activity reflected their depressed mindset.

While posts that were part of the experimental group had discussions revolving around topics like body dysmorphia, drugs, etc, the control group consisted of posts made by healthy individuals discussing a different and a more wider range of topics. For instance, there was in general 4 times more discussion on the topics of "celebrity" in the control group than in the experimental group. Hence, this linguistic analysis of various features clearly demonstrated that there are certain linguistic indicators that can be taken advantage of to successfully classify posts as being made by depressed or healthy individuals.

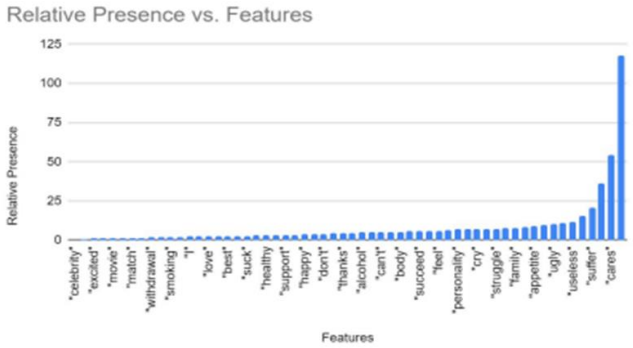


Figure 2

Models

Long Short Term Memory Network and Convolutional Neural Network

A total of three tests were done for this with the control group and each of the subgroups of the experimental group. Test 1 (results in Fig 3.1) was done using the first subgroup containing only tweets, test 2 (results in Fig 3.2) was done using the second subgroup containing both tweets and reddit posts, and test 3 (results in Fig 3.3.1 and 3.3.2) was done using the third subgroup containing only reddit posts. The results for each of the three tests are as follows.

Test three was repeated twice as it rendered the best results. Furthermore, further analysis was also done for Test 3 to prove its accuracy.

Accuracy: 64.22%

	precision	recall	f1-score	support
0	0.71	0.81	0.76	5184
1	0.40	0.27	0.32	2382
accuracy			0.64	7566
macro avg	0.55	0.54	0.54	7566
weighted avg	0.61	0.64	0.62	7566

Figure 3.1: Results for Test 1

Accuracy: 95.00%

	precision	recall	f1-score	support
0	0.97	0.97	0.97	5184
1	0.80	0.80	0.80	732
accuracy			0.95	5916
macro avg	0.88	0.89	0.88	5916
weighted avg	0.95	0.95	0.95	5916

Figure 3.2: Results for Test 2

Accuracy: 88.98%

	precision	recall	f1-score	support
0	0.88	0.98	0.93	5184
1	0.91	0.66	0.76	1930
accuracy			0.89	7114
macro avg	0.90	0.82	0.85	7114
weighted avg	0.89	0.89	0.88	7114

Figure 3.3.1: Results for Test 3 Repeat 1

Accuracy: 95.00%

	precision	recall	f1-score	support
0	0.97	0.97	0.97	5184
1	0.80	0.80	0.80	732
accuracy			0.95	5916
macro avg	0.88	0.89	0.88	5916
weighted avg	0.95	0.95	0.95	5916

Figure 3.3.2: Results for Test 3 Repeat 2

It can clearly be observed that Test 1 gave us an accuracy of 64.22 %, Test 2 gave us an accuracy of 88.98% and both the repeats of Test 3 gave us an accuracy of 95.00%. This discrepancy in the accuracy of the model when different datasets of the experimental group are used can be explained due to the manner in which the experimental group was extracted from twitter. The use of keywords like "Depression" resulted in the API scraping posts that simply consisted of the word "depression", irrespective of whether it was a post by someone truly depressed or by someone healthy. Hence, this skewed and biased dataset could have resulted in the ML Model interpreting a text being significant of depression, when the word "depression" itself was used in the post. Hence, every post part of the control group which contained the word "depression" might have been incorrectly classified as depressed resulting in a lower accuracy.

This fact is further attested as when twitter and reddit datasets are used in conjunction the overall accuracy of the model (88.98%) is lower than when just reddit dataset (95.00) is used, further suggesting how the skewed experimental group extracted from twitter has reduced the accuracy of the model. Hence, subgroups 1 and 2 can be regarded as anomalies due to biased data extraction methods used and were thus not used in the linguistic analysis or testing of the miscellaneous models. Test 3, however, which has been scraped without any biases clearly has the highest accuracy (95.00%) out of all the other subgroups and hence can be used for linguistic analysis and miscellaneous model testing. This can be attributed to not only the fact that no keywords were used, but also that reddit posts tend to be longer than twitter resulting in greater data for the model to train on.

The results of the model when the third subgroup of the experimental group was used was extremely promising. A high precision score for the control group (not depressed) indicated that there were less false positives, in other words there were less depressed posts incorrectly classified as non depressed. Additionally, a high recall also indicated that there were less false negatives, and hence less non-depressed posts that were incorrectly classified as depressed. The precision and recall values for 1 (depressed) was high too (>=0.80) indicating yet again

that there were less false positives and false negatives for this, which is quintessential. Relatively speaking, the reduced recall and precision of 1 in comparison to 0 can be attributed to the number of samples belonging to each class, i.e depressed or not depressed. The support column, which represents this value, clearly shows that there is a lesser sample (732 samples) for testing from the experimental group in comparison to the number of samples from the control group (5184 samples). Hence, this explains the differences, and thus to reduce this in the future we would like to use a larger database for depressed posts too.

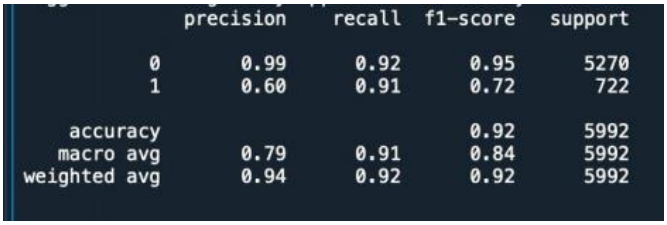
The most important metric too, weighted-average (better than macro average as macro average is a simple arithmetic mean which doesn't take into account the relative number of samples for each class) is 0.95 consistently for recall, precision, and the f-1 score (an harmonious mean of the recall and precision), indicating that apart from the model being accurate, it also has very few false positives and false negatives. Hence, making this 95.00% accurate model extremely fit for forming an early diagnostics system for MDD

Logistic Regression Model

Our logistic regression model was run on subgroup 3 (best dataset as apparent from the discussion carried out previously) of the experimental group and managed to attain an accuracy of 87.627%. While this accuracy is extremely high, we decided to not use this in our app, as our LTSM + CNN Model was much more accurate in predicting and classifying posts according to whether or not the text in the post is indicative of the person being depressed or not.

Multinomial Naive Bayes Classifier Model

Our multinomial naive bayes model was run on subgroup 3 of the experimental group and it managed to attain an accuracy of 92%, as can be seen in Figure 4. The precision, recall, and f-1 score for 0 was extremely high, indicating extremely few false positives and negatives. The precision for 1 was relatively low, due to the fact that the model has given many false positives, or incorrectly classified non-depressed individuals as depressed individuals. Although low precision is generally considered to be unwanted, in the current context it is fine, given that the recall score is high and that it is more important to prevent false negatives (incorrectly classifying depressed individuals as non depressed) than is to prevent false positives. The weighted average precision is 0.94 and the weighted average recall and f1-score both are 0.92, making this model reliable. But, given the LTSM + CNN Model's superiority over this in terms of performance metrics, we decided to use the LTSM + CNN Model in our final app.



	precision	recall	f1-score	support
0	0.99	0.92	0.95	5270
1	0.60	0.91	0.72	722
accuracy			0.92	5992
macro avg	0.79	0.91	0.84	5992
weighted avg	0.94	0.92	0.92	5992

Figure 4: Multinomial Naive Bayes Classifier Model Statistics

App Development

Ultimately, the app could retrieve tweets and display whether the tweets were considered to be depressed or not. To accomplish this, our app registers the user through twitter and then requests to view and analyze the user's current information. After doing so, the machine learning model tests all of their tweets and then the app displays the tweets which have been identified as depressive by the inbuilt LSTM + CNN Model. As seen in figure 5, the tableView which displays the depressive tweets is shown next to the actual tweets made by the user on Twitter. Ultimately, the user can then choose to go to the help page in order to see resources for depression.

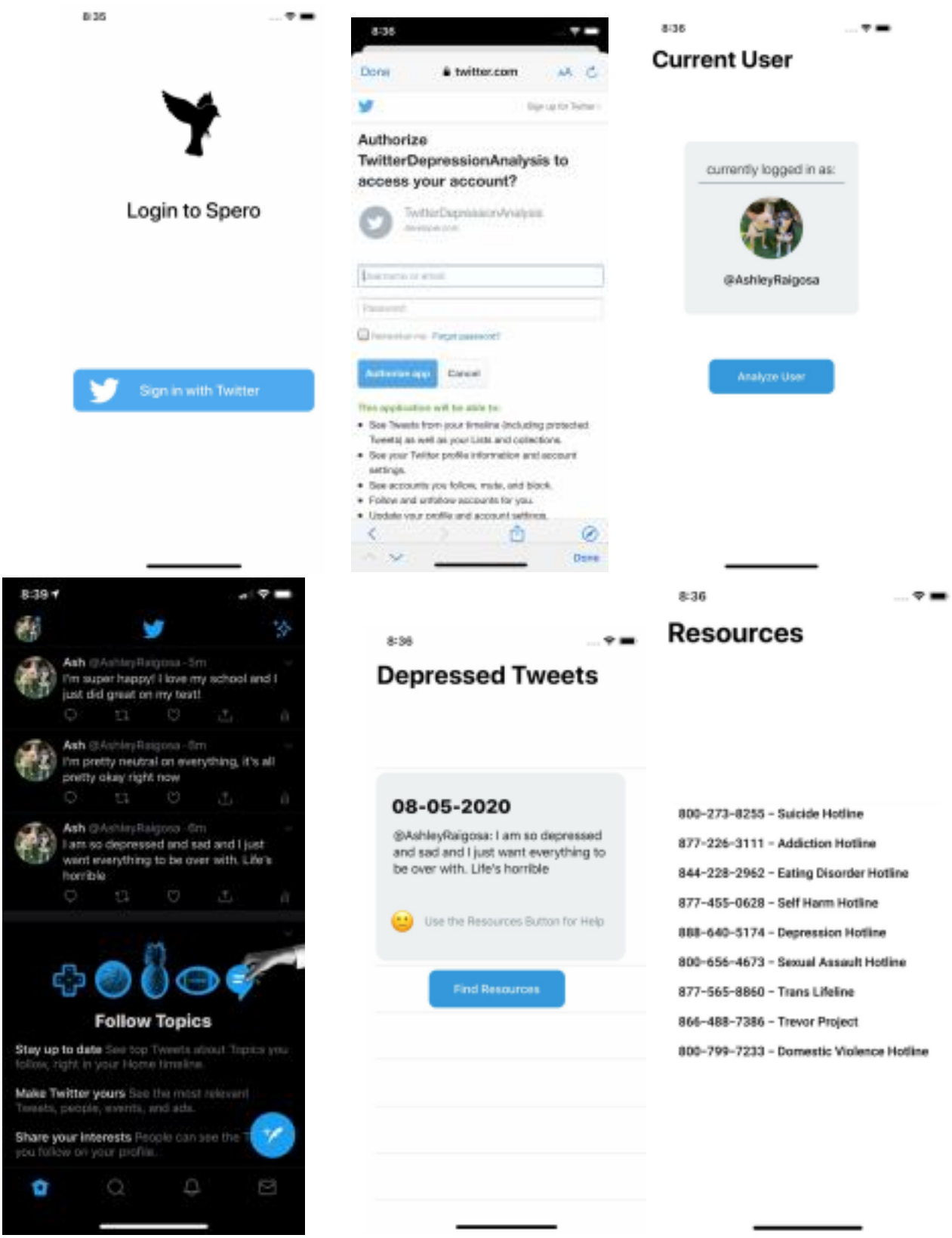

Login to Spero
Sign in with Twitter
twitter.com
Authorize TwitterDepressionAnalysis to access your account?
Current User
currently logged in as:
@AshleyRaigosa
Analyze User
Follow Topics
Depressed Tweets
08-05-2020
@AshleyRaigosa: I am so depressed and sad and I just want everything to be over with. Life's horrible
Find Resources
Resources
800-273-8255 - Suicide Hotline
877-226-3111 - Addiction Hotline
844-228-2962 - Eating Disorder Hotline
877-455-0628 - Self Harm Hotline
888-640-5174 - Depression Hotline
800-656-4673 - Sexual Assault Hotline
877-565-8860 - Trans Lifeline
866-488-7386 - Trevor Project
800-799-7233 - Domestic Violence Hotline

Figure 5

V. CONCLUSION

Through this study, we clearly demonstrated that there exists certain linguistic features that are different amongst depressed and healthy patients. We also built three different ML Algorithms in which LSTM + CNN Model attained an accuracy of 95.00%, Multinomial Naive Bayes Classifier Algorithm attained an accuracy of 92%, and the Logistic Regression Model attained an accuracy of 87.627%. Hence, we chose the LSTM + CNN model, and integrated it into our app system to develop an extremely accurate, 95.00 %, MDD early diagnostics system.

Therefore, we were able to successfully fulfill our targets for this study and we hope that these promising results can help MDD patients across the world.

VI. FUTURE WORK

Further development of the app should be conducted. For instance, an AI Chatbot and a Google API to locate the nearest Psychiatrist should be integrated so that the app could work as a platform and be expandable as more professionals and users join. Additional UI/UX improvements can be made and it can ultimately be released to the general public for use. The navigation between views could be improved and simplifications in the code could be made. Currently, the design of the app is solely a prototype and further work in color pallets and view designs can also be done. Additionally, the app is currently trained on a BOW model to tokenize each tweet to be passed to the machine learning model, as it is the most efficient method to run the model on Swift. But, in the future, instead of using BOW which can be biased, implementing the trained file of Word2Vec or a GloVe would be the most optimal for accuracy. This is because these word vector models will contain all of the trained tokens from the python code of the machine learning model and ensure all data is processed in the same fashion before being completely analyzed by the model. To improve our overall model accuracy, we would also like to use a larger dataset, especially one that has a roughly even distribution of both the classes. Additionally, we would also like to study the potential of other features, such as the time of the day the post has been made, in predicting MDD and ultimately add it as a parameter in the model, if a strong enough association is established with further analysis.

References

VII. REFERENCES

[1]. "Depression", Who.int, 2020. Online. Available: https://www.who.int/news-room/fact-sheets/detail/depression. Accessed: 27- Jul- 2020.

[2]. "Major Depressive Disorder: Symptoms, Causes, and Treatment", Healthline, 2020. Online. Available: https://www.healthline.com/health/clinical-depression. Accessed: 27- Jul- 2020.

[3]. R. Muñoz, P. Mrazek and R. Haggerty, "Institute of Medicine report on prevention of mental disorders: Summary and commentary.", American Psychologist, vol. 51, no. 11, pp. 1116-1122, 1996. Available: 10.1037/0003-066x.51.11.1116.

[4]. A. Halfin, "Depression: The Benefits of Early and Appropriate Treatment", AJMC, 2020. Online. Available: https://www.ajmc.com/journals/supplement/2007/2007-11-vol13-n4suppl/nov07-2638ps092-s097. Accessed: 27- Jul- 2020.

[5]. A. Picardi et al., "A randomised controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care", Journal of Affective Disorders, vol. 198, pp. 96-101, 2016. Available: 10.1016/j.jad.2016.03.025.

[6]. P. Wang et al., "Use of mental health services for anxiety, mood, and substance disorders in 17

countries in the WHO world mental health surveys", The Lancet, vol. 370, no. 9590, pp. 841-850, 2007. Available: 10.1016/s0140-6736(07)61414-7 Accessed 27 July 2020.

[7]. H. M, "Rating depressive patients", PubMed, 2020. Online. Available: https://pubmed.ncbi.nlm.nih.gov/7440521/. Accessed: 27- Jul- 2020.

[8]. I. Cameron et al., "Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II", British Journal of General Practice, vol. 61, no. 588, pp. e419-e426, 2011. Available: 10.3399/bjgp11x583209.

[9]. K. Smarr and A. Keefer, "Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionna", Arthritis Care & Research, vol. 63, no. 11, pp. S454-S466, 2011. Available: 10.1002/acr.20556.

[10]. J. Pennebaker, M. Mehl and K. Niederhoffer, "Psychological Aspects of Natural Language Use: Our Words, Our Selves", Annual Review of Psychology, vol. 54, no. 1, pp. 547-577, 2003. Available: 10.1146/annurev.psych.54.101601.145041.

[11]. V. Prieto, S. Matos, M. Álvarez, F. Cacheda and J. Oliveira, "Twitter: A Good Place to Detect Health Conditions", PLoS ONE, vol. 9, no. 1, p. e86191, 2014. Available: 10.1371/journal.pone.0086191.

[12]. A. Aladağ, S. Muderrisoglu, N. Akbas, O. Zahmacioglu and H. Bingol, "Detecting Suicidal Ideation on Forums: Proof-of-Concept Study", Journal of Medical Internet Research, vol. 20, no. 6, p. e215, 2018. Available: 10.2196/jmir.9840.

[13]. A. Leis, F. Ronzano, M. Mayer, L. Furlong and F. Sanz, "Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis", Journal of Medical Internet Research, vol. 21, no. 6, p. e14199, 2019. Available: 10.2196/14199.

[14]. Y. Tausczik and J. Pennebaker, "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods", Journal of Language and Social Psychology, vol. 29, no. 1, pp. 24-54, 2009. Available: 10.1177/0261927x09351676.

[15]. T. Brockmeyer et al., "Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety", Frontiers in Psychology, vol. 6, 2015. Available: 10.3389/fpsyg.2015.01564.

[16]. E. Seabrook, M. Kern, B. Fulcher and N. Rickard, "Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates", Journal of Medical Internet Research, vol. 20, no. 5, p. e168, 2018. Available: 10.2196/jmir.9267.

[17]. A. Shaw, K. Timpano, T. Tran and J. Joormann, "Correlates of Facebook usage patterns: The relationship between passive Facebook use, social anxiety symptoms, and brooding", Computers in Human Behavior, vol. 48, pp. 575-580, 2015. Available: 10.1016/j.chb.2015.02.003. 18M. Choudhury, M. Gamon, S. Counts and E. Horvitz, "Predicting Depression via Social Media", Microsoft Research, 2020. Online. Available: https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/. Accessed: 27- Jul- 2020.

[18]. "Sentiment140 dataset with 1.6 million tweets", Kaggle.com, 2020. Online. Available: https://www.kaggle.com/kazanova/sentiment140. Accessed: 27- Jul- 2020.

[19]. "What are the DSM-5 criteria for diagnosis of major depressive disorder (clinical depression)?", Medscape.com, 2020. Online. Available: https://www.medscape.com/answers/286759-14692/what-are-the-dsm-5-criteria-for-

diagnosis-of-m ajor-depressive-disorder-clinical-depression. Accessed: 27- Jul- 2020.

[20]."What is Logistic Regression? - Statistics Solutions", Statistics Solutions, 2020. Online. Available: https://www.statisticssolutions.com/what-is-logistic-regression/. Accessed: 28- Jul- 2020. 22H. Lee, B. Tseng, T. Wen and Y. Tsao, "Personalizing Recurrent-Neural-Network-Based Language Model by Social Network", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 3, pp. 519-530, 2017. Available: 10.1109/taslp.2016.2635445.

Cite this article as :

Shivam Garg, Ashley Raigosa, Rimsha Aiman, "Investigating differential linguistic patterns exhibited by Major Depressive Disorder (MDD) Patients and building a Long Short Term Memory Network + Convolutional Neural Network Model, Logistic Regression model, and a Multinomial Naive Bayes Classifier Algorithm to develop Spero, a hybrid app based Early-MDD diagnosis system", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 5, pp. 114-127, September-October 2020. Available at doi : https://doi.org/10.32628/CSEIT206527
Journal URL : http://ijsrcseit.com/CSEIT206527

Appendix

Appendix A: Results of Data Analysis

29	"fat"	170	267	4.352278546	1.003155996	4.338585984
30	"thanks"	474	716	12.13517665	2.690111211	4.511031589
31	"friend"	997	1468	25.52483359	5.515479411	4.627854025
32	"alcohol"	59	81	1.510496672	0.3043282236	4.963380175
33	"never"	962	1312	24.62877624	4.929365795	4.996337717
34	"can't"	615	837	15.74500768	3.144724977	5.006799575
35	"sleep"	399	521	10.21505376	1.957469191	5.218500402
36	"body"	478	623	12.23758321	2.340697325	5.228178404
37	"drugs"	50	62	1.280081925	0.2329425909	5.496267826
38	"succeed"	30	37	0.7680491551	0.1390141268	5.524971977
39	"care"	786	916	20.12288786	3.441538924	5.847060954
40	"feel"	1903	2166	48.71991807	8.137962128	5.98674672
41	"life"	1429	1482	36.58474142	5.568079351	6.57044182
42	"personality"	75	73	1.920122688	0.2742711151	7.000820655
43	"nothing"	598	582	15.30977983	2.186654644	7.001462197
44	"cry"	338	328	8.653353815	1.232341449	7.021880034
45	"die"	718	687	18.38197645	2.581154193	7.121611137
46	"struggle"	157	145	4.019457245	0.5447850917	7.378060279
47	"insomnia"	12	11	0.3072196621	0.0413285242	7.433598659
48	"family"	612	558	15.66820276	2.096483318	7.473564244
49	"fatigue"	6	5	0.153609831	0.01878569282	8.176958525
50	"appetite"	12	9	0.3072196621	0.03381424707	9.085509473
51	"energy"	112	81	2.867383513	0.3043282236	9.422009823
52	"ugly"	90	61	2.304147465	0.2291854524	10.05363753
53	"anxiety"	283	172	7.245263697	0.6462278329	11.21162433
54	"useless"	91	55	2.329749104	0.206642621	11.2742913
55	"trapped"	50	22	1.280081925	0.08265704839	15.48666387
56	"suffer"	267	89	6.835637481	0.3343853321	20.44239631
57	"suicide"	507	95	12.98003072	0.3569281635	36.36594713
58	"cares"	786	98	20.12288786	0.3681995792	54.65212076
59	"likes"	1850	107	47.36303123	0.4020138263	117.8144336

Appendix B: Posts from the Experimental and Control group

Control Group

0		Instead of banning vaping, maybe parent your own kids instead of asking the government to do your job for you.
0		People struggling with money should not have children
0		Any pair of small titties is better than fake pair
0		If Being A Stay-At-Home-Mom Is A Full Time Job Then Being a Working-Dad Is Having 2 Jobs
0		Body cam footage should automatically be backed up to the cloud and accessible by the public.
0		Kobe Bryant's death was tragic but has no right to be next to the worst events of 2020
0		Ice cream in a cup is superior to the cone
0		If you cheated you should have no right to any kind of money/property from your former spouse in a divorce.
0		Society needs to stop defending gang activity and "hood life"
0		All bodies are not beautiful, healthy bodies are beautiful.
0		Men respect women more than women respect men
0		Mental illness is not an excuse for shitty behavior. If you treat others poorly, they have every right to disconnect themselves from you.
0		Farmers are by far the most important job in existence, yet they go under-appreciated, are denounced as 'ruffians' and intellectually challenged by the rest of society.
0		I'm a parent and think schools shouldn't be allowed to give any homework.
0		I think the glorification of single mothers is getting ridiculous, especially if they're single mothers through their own poor life choices.
0		If you think burning the American Flag is free speech, then you should be okay with people flying the confederate flag
0		We should be allowed to see John Cena
0		Please leave your dog at home when going out to restaurants and bars—not everyone likes your dog or wants to be around it
0		It's not just captain marvel, all of the marvel women are written with the exact same personality. they're just bad at writing women, period
0		Who cares who Chick-Fil-A supports? You're going for the chicken, not their values.
0		Adults who repeatedly visit/obsess over Disney are strange
0		I think "gender identity" is the ultimate first world problem
0		Vanilla Coke was the best coke they did
0		Making fun of "edgy teens" is harmful, because we forget their problems are very real to them.
0		Dr. Phil is a garbage person and should not be praised
0		There are too many people who pretend to have a mental illness
0		Colleges applications shouldn't have a race/ ethnicity box
0		If you pay 400+ dollars for a pair of shoes, you're a damn idiot
0		Crapping on kids for not knowing artists who peaked before they were born makes no sense.
0		As a male, I prefer peeing sitting down.
0		Tanning is a "crazy" trend, and pale skin looks so much prettier

Experimental Group