

A Time Series Analysis on Depression During the COVID-19 Period

Alejo Vallega^{#1}, Devin Efendy^{#2}, Kevin Kim^{#3}, Victor Sharonov^{#4}, Joseph Smith^{#5}

[#]*Department of Computer Science, University of Manitoba
66 Chancellors Cir, Canada*

¹vallegaa@myumanitoba.ca

²efendyd@myumanitoba.ca

³kimh3427@myumanitoba.ca

⁴sharonov@myumanitoba.ca

⁵smithj32@myumanitoba.ca

Abstract— This document gives formatting instructions for authors preparing papers for publication in the Proceedings of an IEEE conference. The authors must follow the instructions given in the document for the papers to be published. You can use this document as both an instruction set and as a template into which you can type your own text.

Keywords— COVID-19, Depression, Machine Learning Classifiers, Twitter

I. INTRODUCTION

The novel coronavirus disease (COVID-19) has been declared as a pandemic by the World Health Organization (WHO) on March 11, 2020. It is disheartening that, as of December 2020, around 67 million confirmed cases of COVID-19, including 1.5 million deaths, are reported to WHO [1]. This COVID-19 has continued to be prevalent even in the end of 2020, significantly affecting people's lifestyle and their physical and mental well-being [2].

Depression is one of the common mental disorders found in people during the COVID-19 period. It is a leading cause of disability worldwide and a major contributor to the overall global burden of disease [1]. Traditionally, depression is often diagnosed by a psychological doctor interviewing a patient face-to-face via referring to clinical depression criteria. However people are often ashamed to talk to a psychological doctor or unaware of depression, which leads to more than 70% of people in the early stages of depression not consulting a doctor, letting depression deteriorate

their conditions [3]. Instead of seeking professional assistance, people are increasingly relying on social media platforms, such as Facebook and Twitter, to express their emotions in addition to sharing other information about their lives.

This study exploits data collected from (how many?) user profiles and around 100,000 tweets. Several machine learning classifier techniques are utilized to identify the depression level, which include support vector machines (SVM), Naive Bayes (NB) and Random Forest (RF).

This paper aims to see if there exist new keywords or features from depression-related tweets during the COVID-19 period, and compare them to existing keywords and features that indicate depression of users.

Our key contribution of this paper is a time series of tweets related to depression during the covid-19 period. Specifically, providing a visualization of the number of depression-related tweets on Twitter during the particular COVID-19 period (April to September). The tweets that indicate depression are classified by our Machine Learning model that we trained. In addition to that, after completing classification on tweets using our machine learning models, we are doing feature extraction on the tweets that are classified as depressed during the COVID-19 period.

The rest of the paper is organized as follows: Section II provides background and related works on depression detection. Section III presents which

datasets are used and how datasets are preprocessed prior to applying the machine learning algorithm. Section IV and V describe what features are extracted and which machine learning algorithms are used to detect depression from the datasets. Section VI provides time series analysis (Need more details here). Then, we conclude our study and provide a direction for future work in section VII.

II. BACKGROUND AND RELATED WORK

Efforts to analyze depression of an individual have existed for long before the appearance of the Internet. There are many widely-accepted scales and criteria have been developed based upon the user study or questionnaire survey. In the areas of medicine and psychology, several questionnaire-based measures for rating depression in individuals have been proposed [4, 5, 6, 7]. For instance, Center for Epidemiological Studies-Depression (CES-D), Beck's Depression Scale (BDI) and Zung's Self-rating Depression Scale (SDS) estimate the severity of depression in individuals from the self-reported answers to 20 questions [4, 6, 8]. The questions either have several options aligned with different scores or require participants to evaluate the severity of their circumstances. Then, the level of depression is determined according to the scale of the total score [9]. In general, however, obtaining data through a survey or questionnaire is often expensive and sometimes time-consuming.

Nowadays, it is commonly known that almost everyone has an active account on at least one social media platform, such as Facebook and Twitter, allowing a large amount of data to be generated in a short period of time. For example, Facebook has about 2.7 billion active users and 350 million posts are uploaded in a single day¹. Twitter owns around 340 million active users and 500 million tweets (user posts on Twitter) are generated in a single day, as of December 2020². Through the use of social media's own Application

Programming Interface (API), such as Graph API Explorer (for Facebook) and Twitter Search API, researchers can easily access and obtain large scale data of the users. As the large scale data is available publicly due to the presence of social media, approaches that use such data for depression analysis are seen to be compelling to researchers and thereby provide motivation to analyze the online behaviors of depressed users.

Moreno et al. propose that college students experiencing depressive mood show symptoms consistent with depression on Facebook and place greater investment in social media as a communication outlet because it could be viewed as a safe and indirect outlet for their emotions [10].

Park et al. explored the use of language in describing depressive moods by utilizing real time moods captured from Twitter users and analyzed the differences between Twitter users with and without depression by analyzing their activities [11]. In their later work, a similar analysis is done by analyzing data from Facebook [12].

De Choudhury et al. demonstrate the estimation accuracy that could be achieved by utilizing activities on Twitter to predict depression of the users [13]. They obtained training data for machine learning by crowdsourcing (the practice of engaging a crowd or group for a common goal). Then, models that could be used to predict risk of depression were identified from several features obtained from the records of user activity on Twitter by using Support Vector Machine (SVM). The result of their experiments show that depression can be recognized among Twitter users with an average accuracy of 70% [13]. These approaches are also applied to predict mothers' postpartum depression (the depression after giving birth) from Facebook and Twitter as well [13, 14].

Tsugawa et al. showed that word frequencies are useful for identifying depression and investigated how useful the various features extracted from Twitter user history are for recognizing depression, and the degree of accuracy with which the presence of active depression could be detected by using these features [7, 16].

¹

<https://www.omnicoreagency.com/facebook-statistics/#:~:text=350%20million%20photos%20are%20uploaded,300%2C000%20users%20helping%20in%20translation>

² <https://www.omnicoreagency.com/twitter-statistics/>

Nadeem et al. employed a Bag of Words (BOW) approach which utilizes word occurrence frequencies to quantify the content of a tweet (i.e. putting all words within a bag and measuring how commonly each word appeared). Then they used four types of binary classifiers: linear SVM classifier, decision tree (DT), Naïve Bayes (NB) algorithm, and logistic regression. They found that NB algorithm produced an accuracy of 81% and precision of 0.86, achieving better performance than the other classifiers [15].

Beyond the technical considerations and challenges, there are also ethical considerations to be taken into account when proposing the use of social media data to judge levels of depression. The use of the large volume of data available from social media platforms such as Twitter has potential to aid in early detection of depression [18], however there is still doubt as to the ethically appropriate use of this data [17]. Mikal et. al found that many users of Twitter did not understand the permanence of their posts, as well as either a poor or total lack of understanding of the data tools that could be used to analyze their activity on the platform [17]. There are also concerns about privacy and consent, as often users are not informed that their data can and will be used in this way [19].

In this study we are considering purely the technical aspects of social media data analysis for the purpose of identifying depression. This paper should not be taken to be a statement of our position on the question of the ethics of putting ours or similar methods into practice.

Drawbacks of mentioned works?

Write how **our work** is going to be different than the aforementioned works.

III. DATA COLLECTION

In this study, datasets are collected from IEEE Dataport, submitted by Rabindra, Lamsal. Datasets include CSV files that contain tweet IDs and sentiment scores of the tweets related to the COVID-19 pandemic [20]. For this study, a total of 32 datasets were downloaded out of 276. Each dataset has a one week interval to the preceding and

subsequent dataset. Datasets between March 20 and April 17 contain 1 millions tweets in average. Datasets after April 17 are collected with the addition of some more coronavirus specific keywords, which significantly increased the number of tweets captured in a day from 1 million to 3.5 million tweets. Using a python script, 20,000 random samples of tweet IDs are separated from each dataset to simplify the work. Then, these datasets are hydrated (i.e. extracted) by using Hydrator to extract the real data.

Hydrator is an Electron based desktop application for hydrating Twitter ID datasets. Twitter's Terms of Service do not allow the full JSON for datasets of tweets to be distributed to third parties. However, Twitter does allow datasets consisting of tweet IDs to be shared. Hydrator helps users turn these tweet IDs back into JavaScript Object Notation (JSON) and CSV files on their personal computers prior to processing³.

Preprocessing:

Used python scripts to remove any unnecessary emojis, spaces, articles (a and the) and etc.

([IEEE Covid-19 datasets***](#))

(<https://link.springer.com/article/10.1007/s10489-020-02029-z>) -> Design and analysis of a large-scale COVID-19 tweets dataset.

Did we manually label the data after hydrating the tweet ID?

Talk about why we wanted to exclude/include specific stopwords given how depressed users tend to use more first person pronouns .

Talk about:

³ Documenting the Now. (2020). Hydrator [Computer Software]. Retrieved from <https://github.com/docnow/hydrator>

- [swcwang/depression-detection](#) dataset and how the tweets related to depression are manually picked.
- Also, how we used this dataset for machine learning purposes.
- This dataset did not meet the best accuracy due to lack of samples so we used other multiple datasets but were only extracted using depression hashtag. Which of course lacks reliability.

IV. FEATURE EXTRACTION

For feature extraction, we decided to use the following feature set [21]: Textacy readability index [23]; Textacy text statistic and word count [23]; manually built lexicons which include: suicidality, economic stress, isolation, substance use, domestic stress, guns; sentiment intensity analysis from nltk library; and, tf-idf vector. The differences between our features and the referenced paper is that we don't use the LIWC features set (explained in the limitations section).

For extracting words using tf-idf we selected the top 256 words which occur in at least 2 documents (in our case these are tweets) and at maximum of 80% of the corpus (the whole set of tweets). Among all the tweets there are a lot of common words such as "I", "you", "be", "this". Therefore, for the first time we tried to use tf-idf to extract the features we did not find a lot of words that are interesting because those common words occurred more than the other words. Hence, when calculating tf-idf, we used English stopwords from nltk.corpus to make sure that we extracted meaningful words.

The same tf-idf calculation (occurrence frequency and number of extracted words) is used for both training data set and test data set. To achieve this, we first fit our training data set with TfidfVectorizer from sklearn and save it for later use. Then, we load the same vectorizer that we fit on the training data set to transform the test data set.

V. MACHINE LEARNING ALGORITHM

In this study, three different classifiers are used: Support Vector Model, Naïve Bayes and Random Forest.

A support vector model (SVM) is a supervised learning model that draws a hyperplane in a high-dimensional space to classify two different classes [27]. In this study, we used the linear-kernel SVM.

A Naïve Bayes classifier is a simple learning algorithm that uses Bayes' rule together with a naive assumption that the features used in the algorithm are conditionally independent of each other, which offers computational efficiency [22]. In practice, despite its simplicity, NB is often seen as competitive compared to other classifiers [25, 26]. In this study, we used the Multinomial NB algorithm.

Random Forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [28].

In this study, we used (what kind of RF is used in python?)

Elaborate why an algorithm performed better than the other (As of Dec 20, SVM performed better than NB).

Elaborate on parameter tuning for accuracy improvement

[Mention that our model does not accurately classify depressed tweets but overall the result have negative tone] Also, gives examples of classified tweets

(Talk about the metric results for each algorithm, i.e. a table showing the performance (accuracy on detecting depression from tweets) of each classifiers)

VI. TIME SERIES ANALYSIS AND WORD CLOUD

In this section, we will talk about the result of our trained Machine Learning model. We did the time series analysis from March 20, 2020 to October 09, 2020. For each week during that period, we pick one day to be classified using our trained model

(e.g., March 20, March 27, April 03, etc.). We have a total of 31 dates. For each date, the dataset for that date is about ranging from 1 million to 3 million tweets. Since we don't have the computing power and our API calls are limited, we took 20,000 samples from each date. The 20,000 tweets sample might contain re-tweets of the same thing. Therefore, we remove duplicates re-tweets. So, our sample might be less than 20,000 tweets. Lastly, we run our trained model on those particular dates, which contain sampled tweets, for classification.

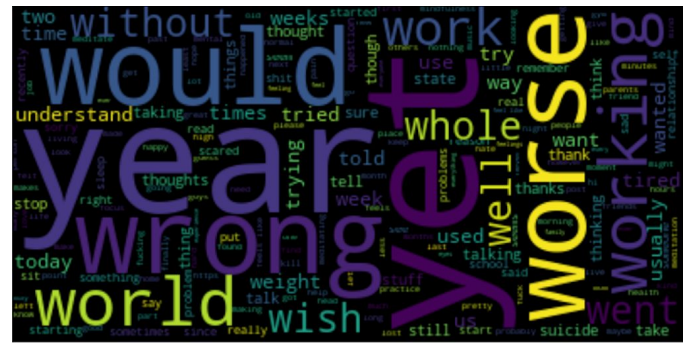


Fig 5.3 (so for this one?)

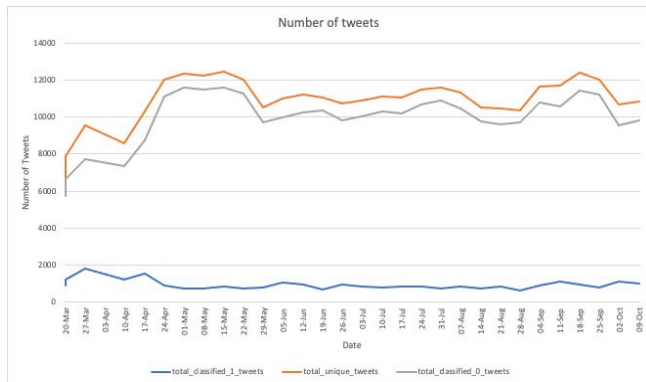


Fig 5.1 (I wonder if we can enlarge the size of this diagram?)

Fig 5.1 explains

([IEEE Covid-19 datasets](#)*** that we used to create the time series analysis)

Talk about results. Evaluation metrics

Any particular differences (or interesting aspects) compared to existing keywords and features that indicate depression of users?

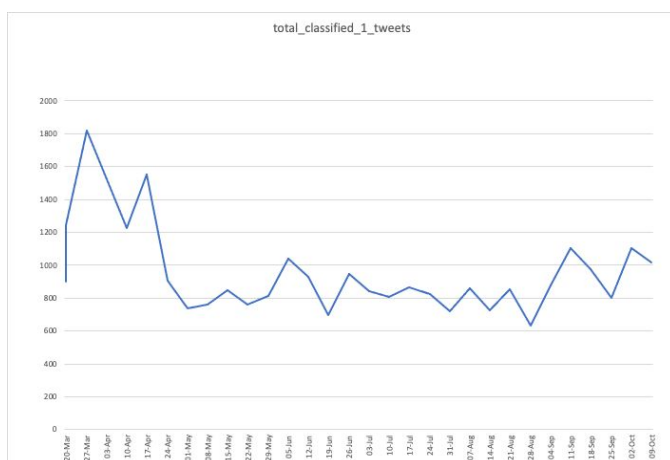


Fig 5.2 (I wonder if we can enlarge the size of this diagram?)

VII. EVALUATION

Talk about results, display diagrams if you have any, evaluation metrics

There is certainly significant room for future development of our methods outlined in this paper to improve accuracy and specificity. As Seabrook et. al explained, while specific keywords and topics in social media posts can be used to indicate depression with a high degree of accuracy, this approach also leads to a high degree of false positives being marked as indicative of depression [24]. They further explain that analyzing the change in emotion across different posts for patterns that indicate depression could provide improvement in specificity of the results [24] to help screen out false positives.

A future version of our algorithm could add additional processing to analyze the changing depression indicators between different tweets from the same person to prune the data of those users whose emotional patterns do not correspond to those considered indicative of depression.

It would also be useful to investigate the posting patterns of users who are depressed as compared to those who are not, such as when they tend to post, how often they post, and what location they generally make posts with different indicators from. If this data has a statistically significant difference between depressed and non-depressed people, it could be used to further prune and refine the results.

Additionally, it would be useful in the future to test our methods on tweets from known individuals,

who could be classified into known depressed and control groups. This would allow us to test the accuracy of our method using posts known by people whose depression status is known, and compare those results to those for which we used indicators suggested by previous works.

VIII. CONCLUSION

In this study, what have tried to do with the datasets we collected?

Make sure to sort the references in lexicographical order before submitting the report! Then, fix the reference numbers used in the report, too.

REFERENCES

- [1] W. H. Organization. (2017). Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization. [Online]. Available: <http://www.who.int/en/newsroom/fact-sheets/detail/depression>
- [2] Bueno-Notivol, Juan et al. "Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies." *International journal of clinical and health psychology : IJCHP*, 100196. 31 Aug. 2020, doi:10.1016/j.ijchp.2020.07.007
- [3] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., & Zhu, W. Depression detection via harvesting social media: A multimodal dictionary learning solution in *IJCAI* 3838-3844 (2017).
- [4] Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. An inventory for measuring depression. *Archives of General Psychiatry* 4, 6 (June 1961), 561–571.
- [5] Hamilton, M. Development of a rating scale for primary depressive illness. *British Journal of Social & Clinical Psychology* 6, 4 (Dec. 1967), 278–296.
- [6] Radloff, L. S. The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1, 3 (1977), 385–401.
- [7] Tsugawa, Kikuchi. "Recognizing Depression from Twitter Activity." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 2015*, pp. 3187–96, doi:10.1145/2702123.2702280.
- [8] Zhung, W. W. K. A self-rating depression scale. *Archives of General Psychiatry* 12, 1 (1965), 63–70.
- [9] Shen, Guangyao, et al. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." *IJCAI*. 2017.
- [10] Moreno, Jelenchick. "Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site." *Depression and Anxiety*, vol. 28, no. 6, Wiley, June 2011, pp. 447–55, doi:10.1002/da.20805.
- [11] Park, M., Cha, C., and Cha, M. Depressive moods of users portrayed in twitter. In *Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD'12)* (Aug. 2012), 1–8.
- [12] Park, S., Lee, S. W., Kwak, J., Cha, M., and Jeong, B. Activities on Facebook reveal the depressive state of users. *Journal of Medical Internet Research* 15, 10 (Oct. 2013), e217.
- [13] De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM'13)* (July 2013), 128–137.
- [14] De Choudhury, M., Counts, S., and Horvitz, E. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)* (Apr. 2013), 3267–3276.
- [15] Nadeem, Moin. Identifying Depression on Twitter. July 2016.
- [16] Tsugawa, S., Mogi, Y., Kikuchi, Y., Kishino, F., Fujita, K., Itoh, Y., and Ohsaki, H. On estimating depressive tendency of twitter users from their tweet data. In *Proceedings of the 2nd International Workshop on Ambient Information Technologies (AMBIT'12)* (Mar. 2013), 29–32.
- [17] Mikal, H. (2016) Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC medical ethics*. [Online] 17 (1), 22–22.
- [18] Reece, R. (2017) Forecasting the onset and course of mental illness with Twitter data. *Scientific reports*. [Online] 7 (1), 13006–13011.
- [19] Nicholas, O. (2020) Ethics and Privacy in Social Media Research for Mental Health. *Current psychiatry reports*. [Online] 22 (12), 84–84.
- [20] Rabindra Lamsal, March 13, 2020, "Coronavirus (COVID-19) Tweets Dataset", IEEE Dataport, doi: <https://dx.doi.org/10.21227/781w-ef42>.
- [21] Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020, July 13). Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: an observational study. <https://doi.org/10.31234/osf.io/xvwcy>
- [22] Webb, Geoffrey I., Eamonn Keogh, and Risto Miikkulainen. "Naïve Bayes." *Encyclopedia of machine learning* 15 (2010): 713-714.
- [23] DeWilde B. textacy Documentation. 2020. Available from: <https://buildmedia.readthedocs.org/media/pdf/textacy/latest/textacy.pdf>
- [24] Seabrook, K. (2018) Predicting Depression From Language-Based Emotion Dynamics: Longitudinal Analysis of Facebook and Twitter Status Updates. *Journal of medical Internet research*. [Online] 20 (5), e168–e168.
- [25] Rish, Irina. "An empirical study of the naive Bayes classifier." *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22. 2001.
- [26] Zhang, H.. "The Optimality of Naive Bayes." *FLAIRS Conference* (2004).
- [27] Noble, William. "What Is a Support Vector Machine?" *Nature Biotechnology*, vol. 24, no. 12, Springer Science and Business Media LLC, Dec. 2006, pp. 1565–67, doi:10.1038/nbt1206-1565.
- [28] Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>

- [29] Write Reference Here
- [30] Write Reference Here
- [31] Write Reference Here
- [32] Write Reference Here
- [33] Write Reference Here
- [34] Write Reference Here
- [35] Write Reference Here
- [36] Write Reference Here
- [37] Write Reference Here
- [38] Write Reference Here
- [39] Write Reference Here
- [40] Write Reference Here
- [41] Write Reference Here
- [42] Write Reference Here
- [43] Write Reference Here
- [44] Write Reference Here

Footer (Click at the sentence where you want to add
footnote -> Insert -> Footnote)

- 1) <https://www.omnicoreagency.com/facebook-statistics/#:~:text=350%20million%20photos%20are%20uploaded,300%2C000%20users%20helping%20in%20translation>
- 2) <https://www.omnicoreagency.com/twitter-statistics/>
- 3) Documenting the Now. (2020). Hydrator [Computer Software]. Retrieved from <https://github.com/docnow/hydrator>