**Predicting Suicidal Patients during COVID-19**

|  |  |
| --- | --- |
| **FNU Abdullah** | **Himani Shah** |
| Luddy School of Informatics | Luddy School of Informatics |
| Indiana University | Indiana University |
| Bloomington, IN 47403 | Bloomington, IN 47403 |
| abdullaf@iu.edu | shahhi@iu.edu |
|  |  |

|  |  |
| --- | --- |
|  |  |

1. Introduction

COVID-19 has had a major impact on all lives, both physically and mentally, people lost their loves ones, suffered losses through their businesses and unemployment. Many of the people, who survived through this pandemic, were affected, developed depression and suicidal tendencies. This issue is further exacerbated because many people bottle up their feelings instead of opening up.

Mental health is a formidable worldwide challenge. In economic terms, combining direct and indirect costs, the global cost of mental health conditions for 2010 was estimated at $2.5 trillion dollars and is expected to grow to $6 trillion by 2030 [Bloom et al. 2021]. In the United States, suicide is the second leading cause of death among those aged 10-34 and the fourth among those aged 35-54, and worldwide 800,000 people are lost to suicide each year [WHO 2014]. Access to professional help is inadequate – in the United States, more than 120 million people live in

federally designated Mental Health Care Professional Health Professional Shortage Areas [HRSA 2021]. This number has increased due to the COVID-19 pandemic (Laura Biester et al. 2020). Since people are more likely to express their emotions to an online community than offline due to the certain degree of anonymity, thus, as a result, we, as researchers, felt the need to identify suicidal people through the social media, especially reddit. While researchers have already analyzed the effects of covid-19 on mental health (Laura Biester et al. 2020) on various subreddits, work has not been done to classify these people into whether they are suicidal or not. For our experiments, we seek to refer to a subreddit (r/suicidewatch) for the purpose of building our models and then run our tests on a test set derived from the subreddits like r/depression and r/anxiety. This way we can, to an extent, identify suicidal people and possibly devise a bot to reach out to such people through a direct message.

1. Related Work

[3]In 2015, an estimated 42,000 Americans died by suicide, each one leaving a lasting impact on friends and family. Because there is a scarcity of data and information regarding those who attempt suicide, scientific research has been limited. They look at data from Twitter users who have tried suicide and do an exploratory study of language and emotion patterns around their attempt. They also compare people who have tried suicide with matched controls. They identify measurable suicide attempt signals in the language of social media data and use these signals to quantify the performance of a basic machine learning classifier as a nonlinear regression model.

[4] Current approaches for assessing population-level mental health necessitate the expensive collecting of large samples of data via instruments like surveys, and so are slow to reflect current, fast changing social situations. This limits the ease with which population-level mental health data may be used to inform health and policy decisions. They show that using natural language processing to analyze publicly available social media data can yield real-time estimates of psychological distress in the population (particularly, English-speaking Twitter users in the United States). They look at changes in language correlates of mental health symptoms at the population level in reaction to the COVID-19 outbreak and George Floyd's death. They compare social media data from healthcare providers to a control group as a case study. Their findings show how computational social science tools can be used to provide real-time or near-real-time insights into the influence of public events on mental health.

[5] They show that social media data can be used to identify people who are at danger of suicide. They employ natural language processing and machine learning (particularly deep learning) approaches to detect measurable signals around suicide attempts and describe designs for an automated system for evaluating suicide risk that can be used by people who don't have an expertise in mental health (e.g., a primary care doctor). They also look at the ethical concerns of using such technology and the privacy consequences. Currently, this technology is only utilized for intervention for people who have "opted in" for the analysis and intervention, but it allows for scalable suicide risk screening, potentially identifying many people who are at risk before they engage with the health care system. This presents a significant cultural challenge about the trade-off between privacy and prevention—they have potentially life-saving technology that is only reaching a fraction of the people who could be at risk due to privacy concerns.

[6] Using social media post data provided in (Macavaney et al., 2021) via the CLPsych 2021 shared task, they propose a deep learning architecture and test three alternative machine learning models to automatically detect persons who may attempt suicide within (1) 30 days and (2) six months. In addition, they develop and extract three sets of handcrafted features for detecting suicide risk, based on the three-stage theory of suicide and previous research on emotions and pronoun use in people with suicidal ideations. Extensive testing reveals that several classic machine learning approaches exceed the baseline on subtask 1 with an F1 score of 0.741 and an F2 score of 0.833. (Prediction of a suicide attempt 30 days prior). On subtask 2 (prediction of suicide 6 months previously), however, the suggested deep learning system surpasses the baseline with an F1 score of 0.737 and an F2 score of 0.843.

[7] In order to quantify the effect of COVID19 on mental health, this pilot investigation employs a deep LSTM neural network with fastText embeddings to predict population rates of depression on Reddit. They discovered that depression rates on Reddit have increased by 50% year over year, implying a 15-million-person rise in depressed Americans and a $7.5 billion increase in depression-related spending. This discovery comes at a time when there is still a lot of doubt about the influence of COVID-19 on physical and economic health, and it shows that mental health should be considered as well. Further research will be required as more data becomes available to validate the findings of this exploratory analysis.

[8] Like many other disease outbreaks before it, the COVID-19 pandemic is anticipated to have a significant impact on mental health. Understanding how it affects people can help you devise tactics for avoiding bad repercussions. By researching talks across mental health support networks on Reddit, they want to gain a better understanding of the consequences of COVID-19 on mental health.First, they quantify the rate at which COVID-19 is discussed in each community, or subreddit, in order to understand levels of pandemic-related discussion. Next, they examine the volume of activity in order to determine whether the number of people discussing mental health has risen. Finally, they analyze how COVID-19 has influenced language use and topics of discussion within each subreddit.

[9] For the research of suicidal ideation and the evaluation of suicide risk, social media has shown to be a significant resource. Due to its anonymity and focus on topic-based communities (subreddits) that can be indicative of someone's state of mind or interest related mental health illnesses, such as r/SuicideWatch, r/Anxiety, and r/depression, Reddit has emerged as the most promising social media platform. The tiny amount of labeled data has been a difficulty for prior work on suicide risk assessment. They offer empirical research into multiple kinds of weakly-supervised techniques and show that utilizing pseudo-labeling based on related mental health concerns (e.g., anxiety, depression) improves model performance for predicting suicide risk.

1. Data

We collect reddit posts from r/suicidewatch and r/happy as we observed that the people in these subreddits display emotions which are consistent with the name of the subreddits. Additionally, we chose r/happy is because we found the posts there from people who are expressing only happiness and the related emotions and as a result not much data cleaning was necessary.

Chart, bar chart

Description automatically generated

Figure 3.1 Understanding our data

We used the Pushshift API and the corresponding library PMAW to extract 1000 comments & posts from both the subreddits (r/suicidewatch & r/happy) between the dates 1/30/2020 and 9/1/2021 using Python, making use of the inbuilt parameters to query our results and further refine our data collection. Initially, we experimented with PRAW but we realized PMAW works much better in gathering data and there were resources that we could refer as we worked on collecting our dataset. Since, our primary concern was the body of the post, we dropped all other columns. Since we were designing this as a binary classification problem, we had to manually assign a label of suicidal/not-suicidal for each record in the dataset, by reading each post and drawing a conclusion based on our rational thinking, we also removed posts which mentioned a third person being suicidal or any general advice that people leave and this is the reason we collected only 1000 comments because there was immense manual labor required.

It was much easier to label all the posts from r/happy than in r/suicidal because of the aforementioned reasons. We also dropped comments which were removed or deleted as they hold no value and offer no insight.As an auxiliary to our primary datasets, we also collected a third and fourth dataset from the subreddit r/depression and r/ to test and verify the results of our model.

1. Methods

As with any other machine learning problem, most of our time was spent in cleaning and labelling the dataset. After a thorough cleaning of the datasets, we combined the data from r/suicidewatch and r/happy into a single dataset to use for our further analysis and modelling. Since, we had only columns in our datasets, selfText and Suicidal, we were limited by the type of analysis we could do. However, prior to cleaning the dataset, we did perform a superficial analysis that seemed fit for future work. Some of the users seemed to have posted twice in the subreddit, with the maximum being 3, this might indicate that the person who posted this is highly suicidal or recovering or just wants someone to talk to and that their post is not getting enough reaction, or they might be simply attention seeking. From one user who posted twice, we saw that the are simply looking for some options to pull themselves out of depression and avoid the suicidal thoughts. However, we were not able to find these patterns for all these users as this was not the primary goal of this paper. This methodology can be used for future works wherein we can design this problem as a multinomial classification problem with labels such as Highly Suicidal, Suicidal and Not Suicidal.

Chart, histogram

Description automatically generated

Figure 4.1 Sentiment analysis is a very common natural language processing task in which we determine if the text is positive, negative or neutral.

With our cleaned dataset, we plotted a barplot to visualize the data distribution for the number of suicidal/non-suicidal people. We could see that, even though we collected 1000 submissions from each subreddit, we ended up with far less number of people in r/happy than in r/suicidal. Through this observation, we could say that people during COVID were more suicidal than they were happy.

Chart

Description automatically generated

Figure 4.2 Trigrams with maximum frequency

To analyze the emotions of the people, we constructed a wordcloud and we could see the results that one would expect in these two subreddits. People in r/suicide watch exhibited far more negative emotions, while people in r/happy exhibited positive emotions. Interesting to note that, people in both subreddits were either depressed or happy because depending on whether they were in relationships or not.

Chart

Description automatically generated

Figure 4.3 Results of Latent Dirichlet Allocation (LDA) model

Having analyzed these two variables, we then moved on to the next step, modelling. We utilized the sklearn count vectorizer coupled with a wordnetlemmatizer, we followed a neat Kaggle tutorial that further utilized this. This ensured that our words were not repeated and stopwords were also removed. We then split our dataset into train, test and validation set. Our baseline model was a simple logistic regression

1. Evaluation

Our logistic regression model gave us an accuracy of 99.85% and a logloss of 0.213. We could say that such a high training accuracy is probably due to overfitting and the lack of training data but the low log-loss value tells us that the predictions are very close to the truth.

And Naïve Bayes model using bag of words tagging has given accuracy of 81.82 %.

1. Discussion and Conclusion

Through this project we aimed to classify suiicdal people and test the model on different subreddits. We utilized logistic regression and Naive bayes, to model our problem and obtained a decent accuracy, given the time constraint and most importantly, the size of the dataset. When tested against a different subreddit, our model Naïve Bayes performs well with accuracy of 81.82 %.

Our main issue was the size of the dataset, since we collected the dataset from scratch, and labelled and cleaned it manually, we were limited by the manual labour that was required if we were to collect more data. Of course, collecting more data helps in improving the accuracy and the model can generalize much better and give accurate predictions and we aim to propose this for future work. Also, since deep learning models work best on large datasets, we refrained from using any deep learning models. At one point, we thought of collecting more data and deploy the labelling task on a website, where real people will label the dataset manually but this is very time consuming and expensive but given the resources, this certainly is possible. In future, we aim to use such large datasets to run computationally extensive deep learning models like NeuralNets, CNN

1. Bibliography
2. Nazli Goharian, Philip Resnik, Andrew Yates, Molly Ireland, Kate Niederhoffer, Rebecca Resnik “[Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access](https://aclanthology.org/2021.clpsych-1.0.pdf)”, 2021
3. Laura Biester, Katie Matton, Janarthanan Rajendran, Emily Mower Provost, Rada Mihalcea, “Quantifying the Effects of COVID-19 on Mental Health Support Forums”, 2020
4. Coppersmith, G., Ngo, K., Leary, R., & Wood, A. (2016, June). Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the third workshop on computational linguistics and clinical psychology* (pp. 106-117).
5. Fine, A., Crutchley, P., Blase, J., Carroll, J., & Coppersmith, G. (2020, November). Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science* (pp. 50-54).
6. Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, *10*, 1178222618792860.
7. Wang, N., Luo, F., Shivtare, Y., Badal, V. D., Subbalakshmi, K. P., Chandramouli, R., & Lee, E. (2021). Learning Models for Suicide Prediction from Social Media Posts. *arXiv preprint arXiv:2105.03315*.
8. Wolohan, J. T. (2020, July). Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.
9. Biester, L., Matton, K., Rajendran, J., Provost, E. M., & Mihalcea, R. (2020). Quantifying the effects of covid-19 on mental health support forums. *arXiv preprint arXiv:2009.04008*.
10. Yang, C., Zhang, Y., & Muresan, S. (2021). Weakly-Supervised Methods for Suicide Risk Assessment: Role of Related Domains. *arXiv preprint arXiv:2106.02792*.