Detecting and Analyzing Suicidal Posts

Filippo Momentè[1] **Anar Abbas**[2]

[1]filippomomente@kaist.ac.kr
[2]anar.abbasly@kaist.ac.kr

Abstract

This document aims to provide a short presentation of the study we are going to undertake. In the first section, an overview of the problem and analysis of related works is provided along with the goals of this study.
In the second section a brief analysis of the baseline dataset that we are going to enrich with our categories is presented. In the third section, we introduce how we are going to tackle this problem, whereas in the last section we explain how we are going to evaluate our work.

1 Introduction

When an individual chooses to take its own life, it decides that his/her experience in this world is not worth continuing. The pain is so strong that they reach the point where they just want to end it, and they believe death is the only way for this to happen.
This situation has been formalized more precisely by Shneidman. He states that suicide is caused by unbearable psychache, where psychache has been defined as mental pain, or better as an acute state of intense psychological pain associated with feelings of guilt, fear, shame, defeat, humiliation, disgrace, grief, bereftness, dread, woe, loneliness, hopelessness, frustrated love, fractured needs, rage, hostility.(Shneidman, E. S., 1993)(Shneidman, E. S., 1998)
Therefore, psychache should be considered as an important variable hinting the presence of suicide risk.(Ielmini M. et al., 2022)

In our study, we aim to track the reasons behind mental pain by analyzing social media posts, and classifying them by training a Machine Learning model. We aim to classify social media posts into *suicidal, non-suicidal* and further classify suicidal posts into *loss, financial problems, loneliness, bullying, discrimination, abuse, guilt, breakup or fear of loss, shame, other.*

Social networks provide a valuable source of data that can be effectively used for analyzing and detecting suicidal behaviors in the real world. (Jashinsky J. et al., 2013), (Sueki, 2014).
For this reason, various studies attempted to make use of these data in order to study suicidality[1] and suicidal behaviors.
These studies focused mainly on: investigating associations of social media content with suicidal behaviors and suicides occurring in the real world (Sueki, 2014),(Fahey et al., 2020), classifying suicidal tweets based on levels of concern (O'Dea et al., 2015), classifying suicidal posts into categories most likely to beneficially or harmfully affect suicide (Metzler et al., 2021), tracking the Werther effect[2] on social media (Fahey et al., 2018).

Not many studies attempted to use data from social networks in order to investigate the causes that lead individuals to harbour those suicidal feelings. However, the "constellation of suicide-producing needs varies from person to person," and "one has to address the frustrated psychological needs of that particular person"(Shneidman, E. S., 1998). Therefore analyzing posts to understand the suicidal individuals' particular thwarted needs would be meaningful to support and sustain therapy.

2 Dataset

In order to sustain our study, randomly sampled a subset of this[3] dataset, which already divides posts into sucidal and non-suicidal. The source of these posts is Reddit, and our subset contains 400 entries.

[1]https://dictionary.apa.org/suicidality
[2]https://dictionary.apa.org/werther-syndrome
[3]https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch

The new dataset contains a total of 52.5 posts labeled as suicidal, and 47.5 labeled as non-suicidal, which we will categorize further based on the categories presented in the introduction.

3 Proposed Solution

We propose a technique for detecting suicidal profiles and distinguish the reasons behind their psychache. In order to do so, we are going to individuate several features to help the model to distinguish between the different categories precisely. We are going mainly to focus on sentiment analysis[4] and extract expressions/terms associated with each motivation behind those suicidal feelings.
After individuating those features, we are going to train a Random Forest classifier and analyze its performances.

3.1 Pre-processing

The dataset we are going to work on is expected to contain entries that are very noisy. Many posts in the dataset are very highly unstructured, containing special Unicode[5] characters, misspellings, punctuation, and abbreviations, making the learning process difficult for machines. Therefore, for the sake of having a more unclouded content, we are going use techniques like tokenization, stop word removal, stemming and lemmatization. However, some of these pre-processing techniques might result in a loss of insightful information, which must be addressed.

3.2 Feature extraction

Most suicidal users suffer from mental health problems due to psychiatric disorders, financial problems, abuse, loneliness, etc. Therefore, it is common for them to make posts containing particular terms such as killing, alone, death, hate, tired, etc. which are less frequent in non-suicidal posts. We are going to define a list of words, expressions, n-grams and patterns that are more associated with suicidal posts rather than not.
Along with this, we are going to perform sentiment analysis based on emojis.

Nowadays, it has become very typical to express ourselves through emojis online. Sometimes, even one emoji suffices to let others know what we are feeling at that moment. We are going to focus on six major emotions love, joy, surprise, anger, sadness, and fear, based on (Parrott, 2001). As shown in Table 1, some emojis include simple characters that are commonly used in text whereas some include Unicode characters that are hard to explicitly extract. In order to tackle this problem, we are going to exploit some common emojis and smiley faces.

Emojis features	**Examples**
`Love`	:-* :* <3 ♥
`Joy`	xD :-) ☻ :D :o) ☺ :3 :c
`Surprise`	:-O :O :-o : o :-0 8-0
`Anger`	:-J >:(>:O
`Sadness`	:-(:(:’(:’-(
`Fear`	%-) %) v.v

Table 1: Emojis features description.

Sentiment analysis serves our purpose of detecting suicidal posts perfectly as it is directly related to emotions. Various studies also made use of sentiment analysis, like (Pestian, 2012)(Birjali, 2017). However, emoticons might not be sufficient to perform it effectively, therefore we are going to define a large dictionary having positive and negative terms collected from different online resources, and for each post we want to count the number of positive and negative terms, according to the mentioned dictionary.

Finally, in order to distinguish between the different reasons behind those suicidal feelings, we are also going to individuate those patterns, expressions, words and n-grams that are more often associated with a category rather than the others.

4 Evaluation

We are going to evaluate our work by observing the values of accuracy, F-score, precision and recall.

[4] https://en.wikipedia.org/wiki/Sentiment_analysis
[5] https://en.wikipedia.org/wiki/Unicode

References

Marouane Birjali. 2017. Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science*, 113:65–72.

Robert A. Fahey, Jeremy Boo, and Michiko Ueda. 2020. Covariance in diurnal patterns of suicide-related expressions on twitter and recorded suicide deaths. *Social Science Medicine*, 253:112960.

Robert A. Fahey, Tetsuya Matsubayashi, and Michiko Ueda. 2018. Tracking the werther effect on social media: Emotional responses to prominent suicide deaths on twitter and subsequent increases in suicide. *Social Science Medicine*, 219:19–29.

Ielmini M. Lucca G. Trabucchi E. Aspesi G.L. Bellini A. Caselli I. Callegari C. Ielmini M. et al. 2022. Assessing mental pain as a predictive factor of suicide risk in a clinical sample of patients with psychiatric disorders. *Behav. Sci.*, 12:111.

Jashinsky J. Burton S. Hanson C. West J. Giraud-Carrier C. Barnes M. Argyle T. Jashinsky J. et al. 2013. Tracking suicide risk factors through twitter in the us. *Crisis*, 35(1):pp.51–59.

Hannah Metzler, Hubert Baginski, Thomas Niederkrotenthaler, and David Garcia. 2021. Detecting potentially harmful and protective suicide-related content on twitter: A machine learning approach.

Bridianne O'Dea, Stephen Wan, Philip J. Batterham, Alison L. Calear, Cecile Paris, and Helen Christensen. 2015. Detecting suicidality on twitter. *Internet Interventions*, 2(2):183–188.

Parrott. 2001. Emotions in social psychology : essential readings. *Psychology*.

John P. Pestian. 2012. Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, 5s1:BII.S9042. PMID: 22419877.

Shneidman, E. S. 1993. *Suicide as psychache: A clinical approach to self-destructive behavior*. Aronson, Northvale, N.J: J.

Shneidman, E. S. 1998. Further reflections on suicide and psychache. *Suicide and Life-Threatening Behavior*, 28(3):245–250.

Hajime Sueki. 2014. The association of suicide-related twitter use with suicidal behaviour: A cross-sectional study of young internet users in japan. *Journal of affective disorders*, 170C:155–160.