

Unveiling Online Suicide Behavior: What Can We Learn About Mental Health from Suicide Survivors of Reddit?

Suicide is a growing public health problem in the online community. This post analyzes suicide-themed online communication on the social networking platform Reddit.

Introduction :

The widespread use of these social media platforms creates the possibility to analyze shared content to identify and characterize user behavior.

A recent study found that people were more likely to seek assistance from informal online resources than to seek formal treatment from a specialist.

Reddit offers discussion forums with different topics, where people expose their opinions and knowledge.

Here, we focus on the publications posted on the Reddit platform, belonging to people who have already attempted suicide on the « Survivors » thread.

State of the art :

1. Durkheim project investigated the online activity and shared content of groups of US veterans on social networks to identify signs of malicious behavior.
⇒ They found that those who committed suicide frequently record behaviors that indicate fear, excitement, and delusion with about 65% accuracy.
2. Sueki is studying a group of young Twitter users to assess the relationship between suicide and tweets about suicidal behavior ⇒ identified the most important indicators of future suicide.

Research questions :

1. How do suicidal individuals communicate about the method used during suicide attempt experiences?
2. What are the most popular methods used by suicidal individuals to end their life?
3. How to recognize a comment dealing about suicide method?

Methods :

- Step 1 : Data acquisition :

Dataset : all the comments of the thread 'Survivors'. they have scraped all the comments with Python Reddit API Wrapper. On Reddit, all comments are organized into a large tree of comments.

- Step 2 : Data Preprocessing :

Preprocessing of data to remove HTML tags, spaces, sidebars and carriage returns. Use regular expressions to expand the abbreviations of English words. These steps are necessary to properly mark words and identify negatives. (Python NLTK Wordnet)

-Step 3 : Feature extractions :

define the features used in the classification process :

- Trigrams : The lemma by itself makes little sense, use trigrams to explore the sentences and find duplicate patterns.
- NLTK POS Tags : tokenized and tagged every word with NLTK POS tags.
- Customised POS Tags : manually define this Tags in this table. All Tags are self-explanatory.
- Method list : define a list of 34 suicide method
- POSEMO list : contains regular expression to detect emotions of people at the social networks...

-Step 4 : Classification :

Application of supervised machine learning algorithms to automatically predict and extract comment-level suicide methods : the task is a multi-label classification task (bc more labels can be used). Each comment containing at least one annotation was extracted.

Use of 4 models :SVM, Logistic Regression, SGD and Perceptron.

Results :

- POS Tagged Sentence Sequence: We tagged all the sentences with the Customised POS Tags mentioned above and look for regularities ➔ Individual POS tag sequences are very useful for identifying and extracting patterns that characterize how people report methods used in suicide attempts.
- Method Labelling : some methods of suicide are used much more often than others. drug overdose is the most use.
- Comment classification : SVC best classification.

Discussion :

analysis reveals that most of these causes appear in sentences as a negative word. Such signs could be used as an important signal to prevent suicide.

In result, the distribution of suicide method labels shows that drug overdose. Some suggestions would be to identify the most common drugs used to commit suicide and to prescribe them to individuals.