**Detecting Suicidal Posts**

Team Members: -

* Achyutanand Mishra (11542607)
* Manohar Kowthavarapu (11554625)
* Rama Mamidi (11551417)
* NaveenRaja Yalagandula (11521505)

# Goals and Objectives: -

## Motivation: -

It is important to realize that suicide is an operation that starts with feelings of pain, Anxiety and discouragement, progresses through suicidal ideology and planning, and culminates in actual suicide attempts. The importance of integrating psychopathological diagnoses and other variables into a more medically founded approach to suicidal behavior has been demonstrated in several recent studies. Clinicians in the challenging field of preventing teenagers' suicide gestures might benefit from an effective guidance from a qualitative examination of the motives behind the suicidal process. There may not be a significant amount of accurate data in this field because most instruments used to study the suicidal motive are self-report measures.

## Significance: -

## Objective: -

Based on the titles and content of forum posts, we plan to create a classifier that can distinguish between suicidal and non-suicidal messages. The algorithm will analyze a dataset to determine if the title and body content of a particular post include suicidal thoughts. The goal of the performance is to emulate a human expert's predictive accuracy. We train a set of data against psychological variables that are linked to suicide, such as burden, stress, Anxiety.

## Features: -

The method of using natural language processing to analyze text data from two support groups and discover the terms that are used in a virtual suicide note in order to determine the suicidal consequences from the suicidal posts. We want to precisely identify people at risk of suicide using Machine Learning classifiers. We use some methods like data collection, cleaning, preprocessing, Exploratory data analysis to prepare the data for the output. Co-occurrences, TF-IDF, Linguistic approaches, Machine learning approaches to get the output.

## Background: -

Doctors and patients typically engage while assessing the danger of suicide. However, a sizable portion of those who suffer from mental disorders do not obtain treatment because there is a lack of access to psychological care facilities, a shortage of physicians, a lack of awareness, and stigma, neglect, and discrimination associated with mental disorders. In contrast, the use of social media and internet connectivity has significantly increased, offering professionals and sufferers a route for contact that might aid in the creation of techniques for identifying mental health problems among social media users.

## Analysis: -

### Data collection: -

Data collection using Reddit's API was the beginning of our adventure which only allows us to get some unique posts per subreddit.

Graphical user interface, text

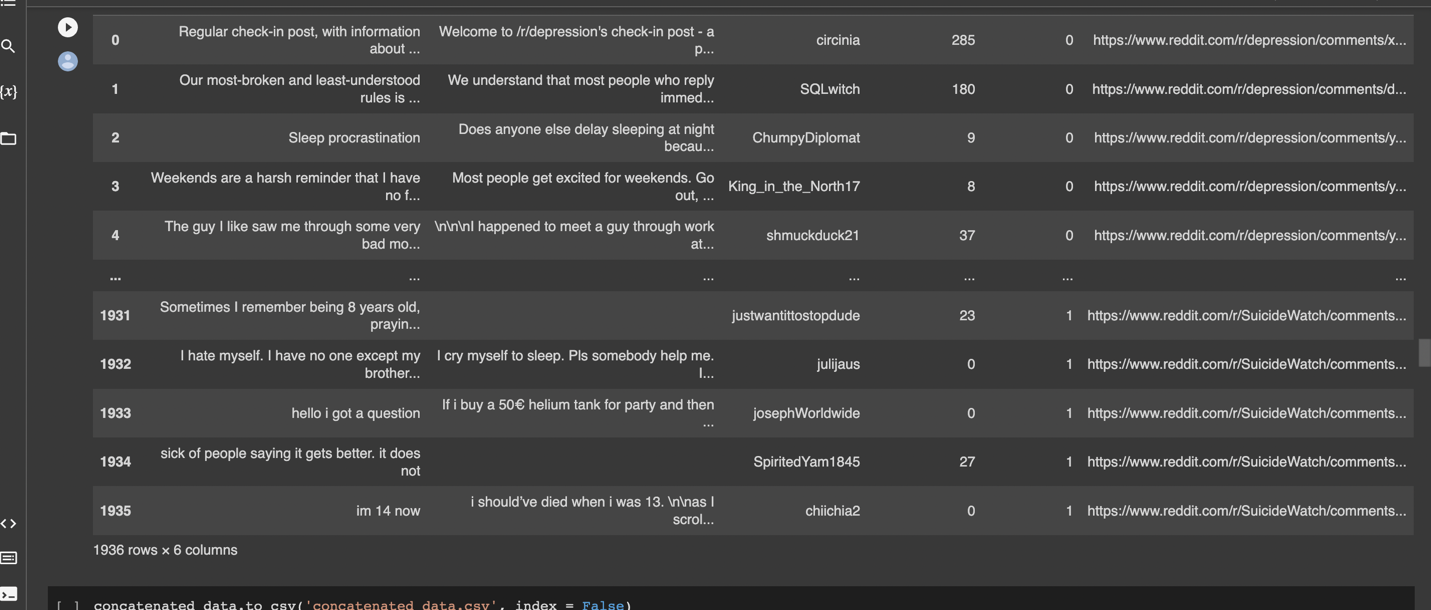
Description automatically generated

### Data cleaning: -

he next thing we did was tidy up our Dataset and account for any missing information. Users who entered nothing in the post field and just entered text in the title field were frequently seen on r/SuicideWatch.

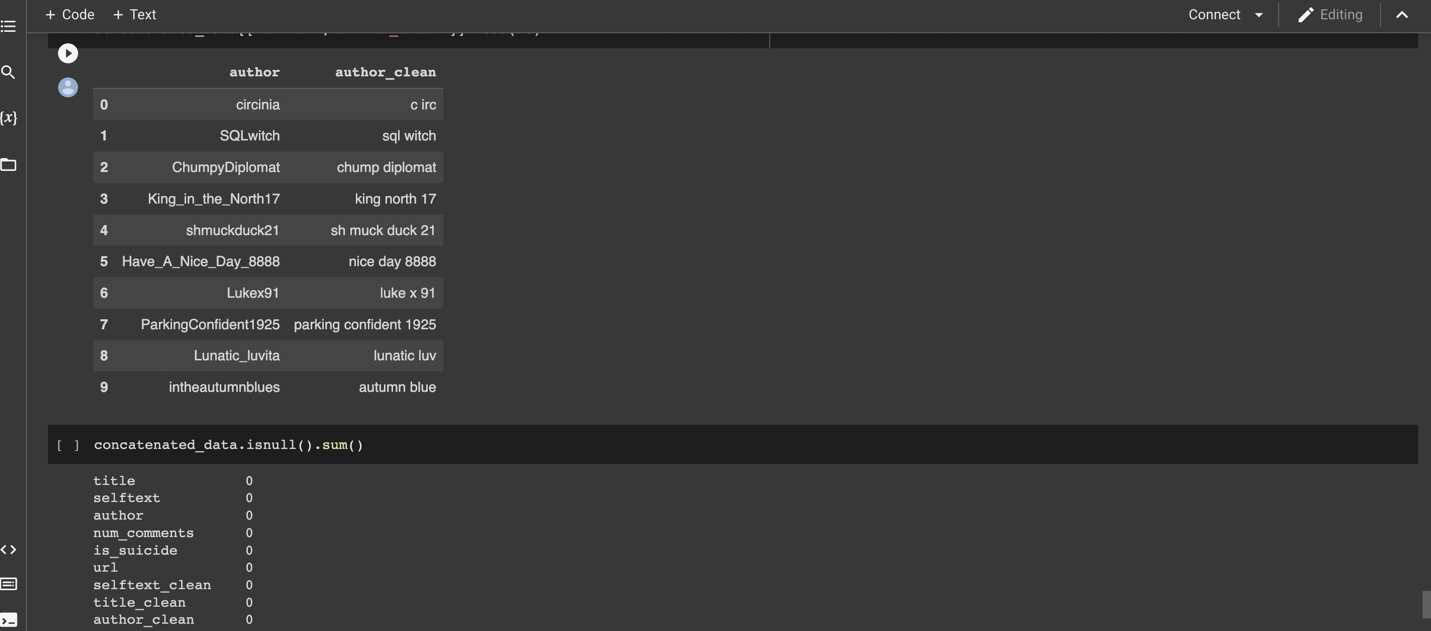
Graphical user interface, text

Description automatically generated



### Data preprocessing: -

we built a pre-processor for removing punctuation, non-alpha characters, lowercase letters, tokenizing, and lemmatizing it using a variety of Python and NLP packages. Stop words were also eliminated.



## Libraries: -

* NumPy
* SciPy
* matplotlib
* Pandas
* Seaborn
* Scikit-Learn
* NLTK
* wordcloud
* spaCy
* TextBlob
* Gensim

## IMPLEMENTATION: -

* Reddit's API used for this project to gather postings from the "r/depression" and "r/SuicideWatch" subreddits.
* To ensure consistency on the data collecting aspect, we try to automate as much of this operation as we can into tidy functions.
* Once we acquire the data, we will examine it for the first time, looking for any null values and deciding which portions of the data set are most helpful for our classifier. We have also started using natural language processing techniques to pre-process the text data.
* To create the model, we'll use a machine learning method like the Naive Bayes theorem to assess the data's or context's accuracy.
* A pipeline for scoring various classifier models, such as Multinomial Naive Bayes and K-Nearest Neighbors, before deciding on the final production model.

## Project management: -

People who have suicide thoughts frequently post their opinions and ideas on social media. As a result, research from multiple studies has shown that it is possible to identify those who are thinking about suicide by looking at social media posts. However, identifying and understanding patterns of suicidal ideation is a difficult undertaking. To automate the early diagnosis of suicidal thoughts, a machine learning system should be created. We make use of Reddit datasets, text representation techniques like TF-IDF and Word2Vec, and machine learning methods for identification.

## Implementation status report: -

### Work completed: -

We have collected the data from the dataset and ran successfully after that we have performed some data cleaning methods to clean the unwanted data from the dataset and later we have done the preprocessing methods like tokenization, lemmatization, stop words etc.

### Responsibility: -

* NaveenRaja Yalagandula (11521505) : - Documentation and helped to get the data. (15% contribution till yet)
* Rama Mamidi (11551417) : - Data collection (25% contribution)
* Manohar Kowthavarapu (11554625) : - Data cleaning (30% contribution)
* Achyutanand Mishra (11542607): - Data preprocessing (30% contribution)

## Work to be completed: -

We are doing work on the four different machine algorithm and it is divided into each member and whatever algorithm gives the best accuracy for this project we can build our model with that algorithm.

### Responsibility: -

As we mentioned Each person have responsibility to perform the machine learning algorithm to get the best accuracy.

References: -

<https://www.linkedin.com/pulse/suicide-ideation-nlp-analysis-sergey-sundukovskiy/>

<https://medium.com/swlh/detecting-depression-in-social-media-via-twitter-usage-2d8f3df9b313>

## Dataset: - <https://www.reddit.com/r/depression.json>

GitHub links: -

<https://github.com/Achyut2995/StackOverflowKeywordidentification> <https://github.com/mkowthavarapu>  
<https://github.com/Srujana390> <https://github.com/NaveenrajaYalagandula/NaveenrajaYalagandula.git>