

Project Description (Team 32)

- **Overview of an industry, business or problem**

According to WHO, over 700,000 people take their lives each year and unfortunately even more individuals attempt suicide. Research suggests that approximately 80%-90% of suicidal people signaled their intentions ([National Library of Medicine](#)). In addition, in 2020 alone, suicide was the 12th leading cause of death in the USA, surpassing the rate of homicides by nearly two times ([National Institute of Mental Health](#)).

- **Define the specific problem that should be solved**

Social media is a platform where millions of individuals take solace in expressing their feelings, emotions, and even their deepest thoughts. Since individuals with suicidal ideation are likely to signal their intentions, there is a fairly high probability that these individuals voice out their feelings or thoughts in some of these platforms, as it has been done many times in the past using historical data. As Twitter is currently one of the most publicly available social media platforms, we plan on using Twitter API to select tweets containing keywords about suicidal ideation and flag "high-risk" suicidal tweets through Natural Language Processing, as a form of prevention or intervention for high-risk profiles and or people with suicidal ideation.

- **Why does this problem matter?**

This problem is essential for detecting high-risk suicidal thoughts/ ideation on social media. As mental health has been an on-going topic over the past few years, especially with the exponential rise of suicidal attempts and depression during covid, it is important to address the matter and figure out a way to reach out to individuals and intervene before suicidal attempts are actually made. Solving this problem has the capability to decrease suicide/suicide attempt rates as measures are taken prior to the attempts.

- **What are the datasets that you will consider to solve this problem?**

Dataset	Description (context and how it was collected)	Major Variables
https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch	The datasets are a collection of post from "SuicideWatch(Dec 16, 2008- Jan 2, 2021)" and "depression(Jan 1, 2009 - Jan 2, 2021)" subreddits of the Reddit platform using Pushshift API	Top words of suicidal texts (tokenized words that people that committed suicide used).

Twitter API	Tweets from people containing keywords indicative of suicidal thoughts or behaviour based on historical data. E.g. suicide, die	Keywords used by people that committed suicide based on historical data
https://crisistrends.org/#visualizations	Data was from over 30 million texts Crisis counseling hotline got from people in crisis. Machine learning was used to pull out words and emojis that can signal a person at higher risk of suicide ideation or self-harm.	Top 35 used words by people having suicidal thoughts. Some of the words are: <ul style="list-style-type: none"> • Suicide • Die • Kill • Life • Bad • Anymore • Never • People • Feeling