Women in the Workforce: Project Scope
DS4A

Team: 32
Members: Jessica Addai, Oluwabunmi Ariyo, Kingsley Yang, Yue Huang, Countess Olufunmi, Helen Kpoto

Problem:

Suicidality involves suicide attempts and ideation. Suicidal ideation can range from having thoughts, ideas, or even ruminations about taking one's own life (Science Direct). According to The World Health Organization (WHO), over 700,000 people take their lives each year and unfortunately even more individuals attempt suicide. In 2020 alone, suicide was the 12th leading cause of death in the USA, surpassing the rate of homicides by nearly two times, with approximately 1.2 million suicide attempts (National Institute of Mental Health). Additionally, it is the second most common cause of death in youths. As social media use has significantly increased over the years, it has also been linked to both depression and suicidality . Research also suggests that self-harming youth are more likely to be most active on these platforms in comparison to those who do not self-harm. In addition, it was discovered that social networking sires were used by both suicidal and self-harming youth as a medium to communicate and seek for support. (National Library of medicine). Furthermore, approximately 80%-90% of suicidal people signaled their intentions (National Library of Medicine). Suicide is a global problem that can easily affect anyone; Therefore, it is imperative to find ways to address the problem as soon signs show before it is too late.

The goal for this problem is to detect 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.

Proposal:

Twitter currently has one of the largest and most publicly available platforms where people are able to voice their thoughts/feelings, opinions, etc. Through the use of Twitter API, we plan on analyzing different tweets through Natural Language Processing (NLP) techniques, with emphasis on sentiment analysis to track and flag suicidal tweets.

Feasibility:
○ *Can your team accomplish this given your current skills and the skills you will be learning throughout the program?*

Yes. Our team will use first use data on Kaggle to explore the words that often associated with suicidal ideation and use these data to train our model. We will then use Twitter API to collect Tweets in the past year and then use them to test our model. Our team has members with backgrounds in data collection, Machine Learning and Natural Language Processing (NLP). For this project, we will use Python as the main tool to conduct our research, and finally we will use PowerPoint to present our results.

○ *Can your team complete this in the time frame of the program?*

We believe we can collect data sources needed by July 20, analyze results by July 23, and finish the project by giving a presentation by July 30. We have regular meetings scheduled for Tuesday afternoon to work before our Saturday deadlines. We plan to communicate regularly over WhatsApp and Zoom and rely on feedback from our TA and mentor.

Impact:

○ *Does the proposal directly address the need(s) identified in the previous step?*

We believe understanding how people express their suicidal ideation is an important way to prevent them to complete suicide. The proposed model will detect the tweets that contain high-risk suicidal thoughts/ideation. We believe if this model is in use, it can help to reduce the suicide rates by communicating with the person in advance and intervening before they actually commit suicide.

○ *Will the final product provide a material improvement over existing solutions/operations in terms of efficiency, accuracy, etc.?*

This topic has been under consideration by the government in health. Some researches have been done so far, including extracting psychiatric stressors for suicide from social media using deep learning. Our approach will build on this information and use Natural Language Processing (NLP) and Machine Learning to provide another way to detect the suicidality on Twitter. Similar research has done before in 2015 (Detecting suicidality on Twitter), and we decide to focus more on the Tweets during Covid and explore the new trend of suicide-related posts.

Usability:
The models and algorithms are used to detect and predict possible suicide and depression by using Twitter to find key words that are markers of suicide and depression. These models can be used to prevent and decrease suicide rates by detecting and predicting possible suicide ideation early enough. The models can be used by any individuals such as parents and any organizations such as health agencies or social agencies so that treatments can be provided to individuals with risk of suicide early enough. We will provide charts and graphs using different data visualization techniques to demonstrate our findings.

Timeline
When do you expect to have each component of your project completed? You can use the due dates in this document as a guideline.

Date	Deliverable	Details
Week 1	Project idea brainstorm	
Week 2	Project detail & outline	
Week 3	Project scoping	**Scoping - Team mid-Week meeting**
Week 4	Data cleaning, aggregation and EDA, Project report draft due	**EDA - Team saturday meeting**
Week 5	Modeling and dashboard building	
Week 6	Project and datafolio	
Week 7	Final report complete and presentation rehearsal	