Youth Mental Health Narratives: Novel Variables
Midpoint Submission - Temporal Extraction

What question, topic, or area related to youth mental health are you planning to explore? How did you decide on this topic? Please mention any references you considered, including research studies, review articles, published theories, etc.

The area that our team is planning to explore is the possible early interventions for youth suicides. However, according to the National Violent Death Reporting System (NVDRS) coding guide, these events are only coded as **boolean variables**. Many studies show that suicide preventions are **time sensitive** - after a known event, there must be targeted and suitable interventions for the individual to successfully prevent escalations.

> "About half of people who die by suicide visit their primary care provider (PCP) within 1 month of doing so, compared with fewer than 1 in 5 contacting specialty mental health."[1]
>
> "...suicide-related outcomes within one-week or one-month in individuals with current suicidal ideation (SI) or a recent suicide attempt (SA)."[2]

The purpose of our exploration is to extract **temporal variables** and tie them to **specific events**. The outcome of the exploration is the ability to plot events leading up to the victim's suicide on a timeline, as well as a timeline for the general population for population health analysis. By analyzing timelines for a population, healthcare authorities may use this information to advise on suitable intervention points from a healthcare system perspective.

The effects of this spread further than just post-mortem reports by law enforcement. We can perform the same analysis on specialty mental health notes, primary care notes, and discharge summaries of hospitals. The hope is that through this exploration we can proactively and effectively intervene on youth suicides through **preventative** measures, rather than reactive.

What methods have you experimented with so far (e.g., data preprocessing, key features, algorithms, other novel aspects of your solution)? Why did you decide to try these methods?

To successfully complete this experiment, we came up with a 4 step process.

1. Sentence Segmentation: This is a preprocessing step to ensure that sentences are being split properly. Although sometimes, sentences are properly split by punctuation, sometimes, it could be missing due to its free-text structure. We employed proper stop word detection to split sentences with best effort.

2. Temporal Variable Extraction: We have trained an NER algorithm capable of recognizing temporal variables using state of the art BERT algorithms. So far, with the assistance of manually annotated data, we are able to obtain a recall of **0.95** and a precision of **0.89**. (See last question for preliminary display of results)

3. Topic Modelling of Events: With millions of permutations on how sentences can be written, and events be depicted, we need to perform topic modeling on currently recognized events that

---

[1] Dueweke, A. R., & Bridges, A. J. (2018). Suicide interventions in primary care: A selective review of the evidence. Families, Systems, & Health, 36(3), 289–302. https://doi.org/10.1037/fsh0000349

[2] Lengvenyte, A., Olié, E., Strumila, R., Navickas, A., Gonzalez Pinto, A., & Courtet, P. (2021). Immediate and short-term efficacy of suicide-targeted interventions in suicidal individuals: A systematic review. The World Journal of Biological Psychiatry, 22(9), 670–685. https://doi.org/10.1080/15622975.2021.1907712

are binary encoded. We perform supervised topic modeling to the existing binary variables. The topics are trained and driven by the existing variables in the NVDRS coding guidelines.

4. Event Reordering & Timeline Reconstruction: We reformat free-text notes into an event log format to perform population process mining. Process mining “is a method of applying specialized algorithms to event log data to identify trends, patterns and details of how a process unfold”[3]. We make use of algorithms such as **Alpha Miner** and **Heuristic Miner** to graph out the possible paths that these victims go through.

What other methods are you planning to test and why?

We want to increase accuracy of temporal variable extraction, and plan on using larger BERT models to increase accuracy. Furthermore, a more sophisticated LLM can help with event sequencing and topic modeling.

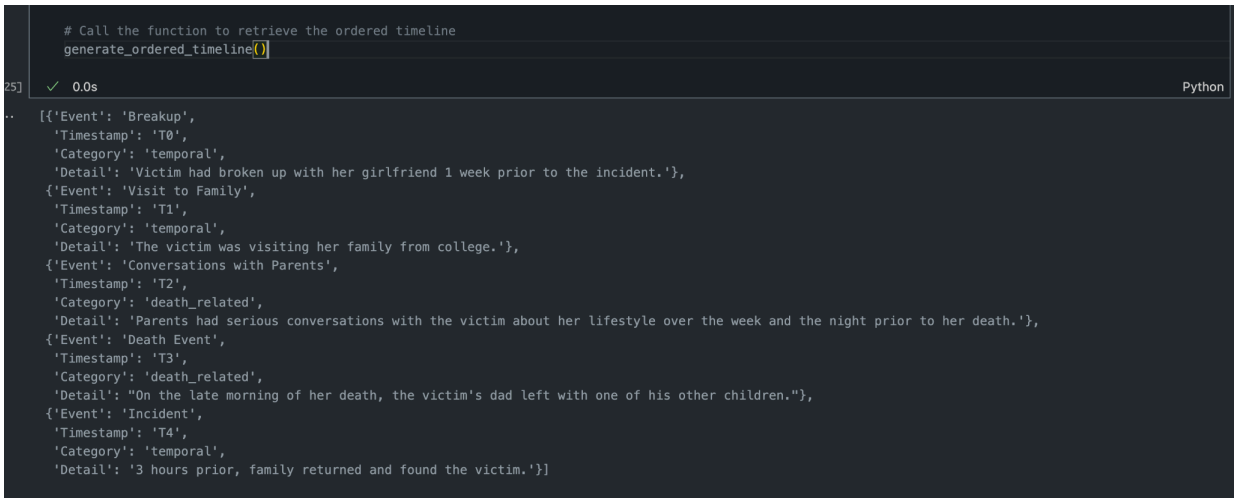
What variables are you planning to explore extracting from the narratives and why? Do you have any findings so far?

We plan on extracting all temporal variables from free-text notes, and correlate that with the events that it refers to.

We also plan on aggregating the results using modern process mining methods like the alpha miner or the heuristic miner.

Do you have any useful tables, charts, graphs, or visualizations from the process so far (e.g., exploring the data, testing different features, summarizing model performance, etc.)?

So far we are able to perform Temporal Variable Extraction and event sequencing for single notes. However, this process is currently very time consuming and requires more optimizations before we can run this with a batch of 4000 in a cost efficient manner.



```
    # Call the function to retrieve the ordered timeline
    generate_ordered_timeline()

25]  ✓ 0.0s                                                                    Python

[{'Event': 'Breakup',
  'Timestamp': 'T0',
  'Category': 'temporal',
  'Detail': 'Victim had broken up with her girlfriend 1 week prior to the incident.'},
 {'Event': 'Visit to Family',
  'Timestamp': 'T1',
  'Category': 'temporal',
  'Detail': 'The victim was visiting her family from college.'},
 {'Event': 'Conversations with Parents',
  'Timestamp': 'T2',
  'Category': 'death_related',
  'Detail': 'Parents had serious conversations with the victim about her lifestyle over the week and the night prior to her death.'},
 {'Event': 'Death Event',
  'Timestamp': 'T3',
  'Category': 'death_related',
  'Detail': "On the late morning of her death, the victim's dad left with one of his other children."},
 {'Event': 'Incident',
  'Timestamp': 'T4',
  'Category': 'temporal',
  'Detail': '3 hours prior, family returned and found the victim.'}]
```

[3] Quote taken from https://www.ibm.com/topics/process-mining