Detecting Suicide Risk and Exploring Contributing Factors: Classification and Topic Modeling of Social Media Data

Evan Dan
Kent State University
Kent, Ohio

Introduction

Issue of Suicide

- Growing health concern globally
- 700,000+ lives each year

Vulnerability of Youth (CDC 2021 survey)

- 29% - poor mental health
- 22% - considered suicide

Traditional Research

- Studies have examined factors influencing suicidal tendencies
- Limitation of traditional methods like surveys
  - Biases and predefined response options
  - incomplete emotions, sentiments, nuances

- Social Media as a Data Source

Background Research

- Advanced Computational Approaches
  - Shortcomings of isolated + static factors
  - Need for multifaceted approaches using dynamic data to analyze interplay
  - Complex models to enhance predictive validity
  - Reddit + ML and NLP techniques
- BERTopic
  - Developed by Grootendorst
  - Tool for topic modeling
  - Embeddings from the LLM BERT to comprehend sentiments
  - BERTopic’s effectiveness in analyzing social media data regarding mental health

The Main Contributions

- Demonstrated value of using Natural Language Processing methodologies, including fine-tuning Llama 3-8b and Mistral-7b, for analyzing social media data regarding suicidal ideation, a topic full of complex nuances.
- The Llama 3-8b model achieved a test accuracy of 0.9371 for classifying Reddit posts for suicidal ideation, demonstrating its ability to capture detailed emotional patterns.
- Using BERTopic, we revealed key topics in the discussions within the classified suicidal and non-suicidal Reddit posts.

Methods

- **Dataset:**
    - Kaggle, 200,000 posts,
    - r/SuicideWatch and r/Teenagers
- Data **Pre-processing:**
    - Remove noise
    - Standardize formatting
- Llama 3-8b
    - LLM with 8B parameters
    - Pretrained on ~15 trillion tokens of publicly available data
- Mistral 7B
    - Pre-trained LLM with 7.3 billion parameters
    - Identify complex relationships

Methods-Classification

Manual Labeling:

- 900 posts hand-labeled as “suicidal” or “not suicidal” according DSM-5
- Split into 80% for training, 10% for validation, and 10% for testing

Training:

- 3 epochs with a learning rate 0.0002

- **Evaluation Metrics**
  - Accuracy, precision, recall, and F-1 scores

- Re-label entire dataset

Methods-Topic Modeling

BERTopic:

- Comprehensive and in-depth analysis
- Extracts discussion topics – understand underlying themes

KeyBERTInspired:

- Reduce stopwords for precise, meaningful topics

- **OpenAI's GPT-4o:**
  - Create accurate labels for the topics using this prompt:

I have a topic that contains the following documents: [DOCUMENTS]
The topic is described by the following keywords: [KEYWORDS].
On the basis of the information above, extract a short but highly descriptive topic label. Make sure it is in the following format: topic: <topic label>.

Methods-Topic Modeling

Intertopic Distance Maps: :

- Visual representations of spatial relationships and groupings of the topics in a 2D plane

Similarity Matrices:

- Cosine similarity between the topic embeddings
- Analysis of the similarities between each pair of topics
- identify closely related topics

Results-Classification Model

- Llama 3-8b had better test performance
- Relabeled entire dataset using the fine-tuned Llama 3-8B model

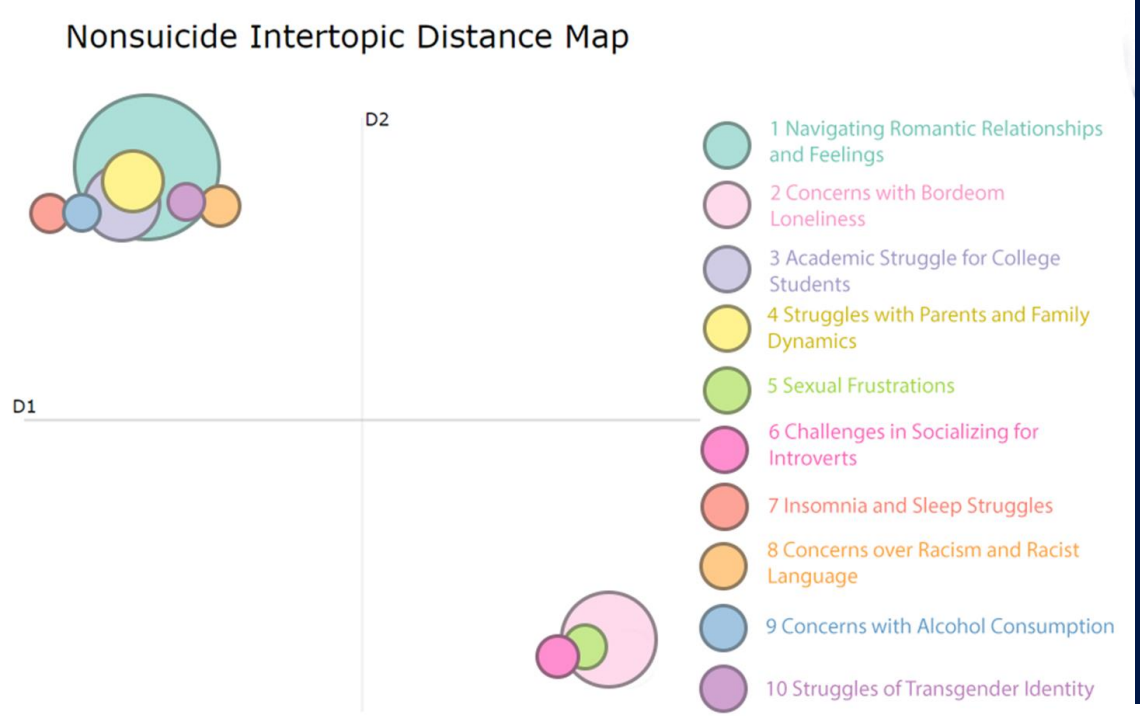
Table II Model Performance Comparison

Model	Test Performance			
	Accuracy	Recall	Precision	F1 Score
Llama 3-8b	0.9371	0.9371	1.000	0.9676
Mistral-7b	0.9314	0.9314	1.000	0.9645

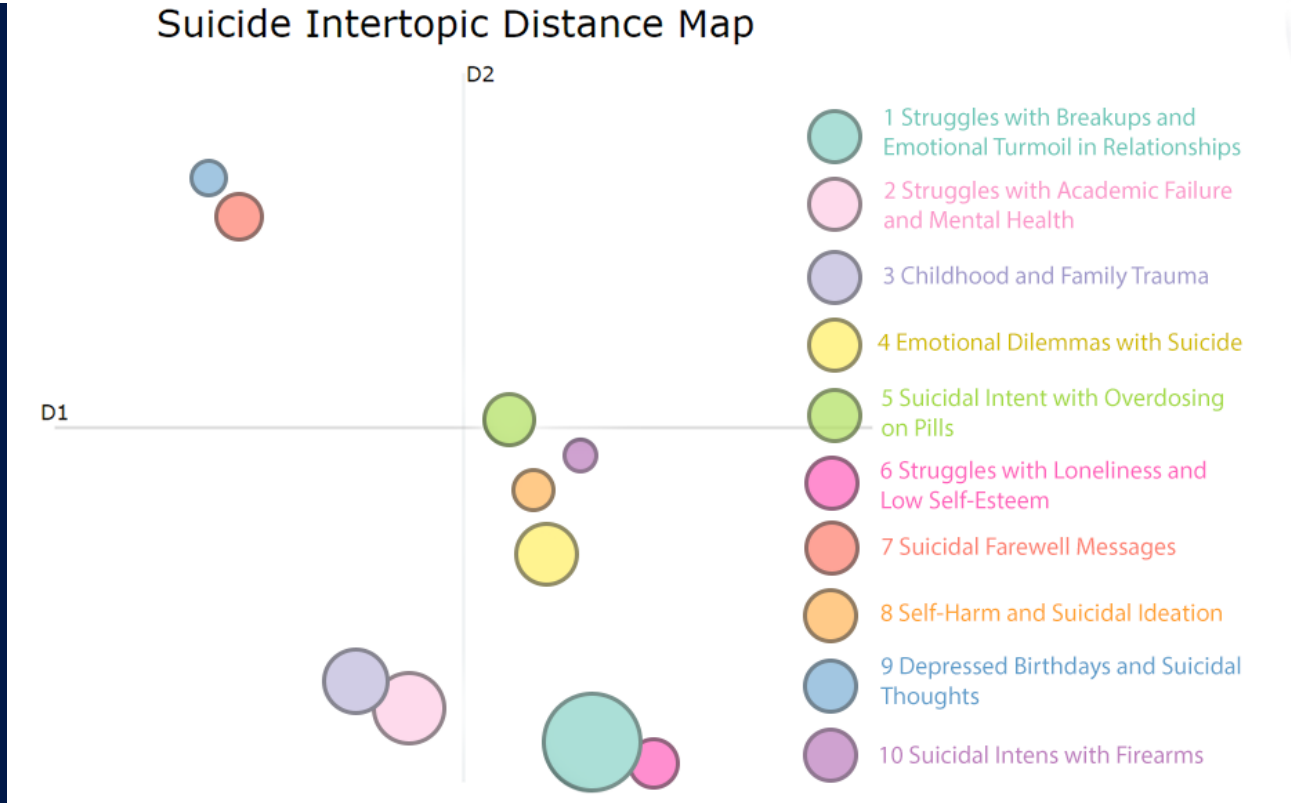
Results-Key Topics

Table III Top 10 Topics in Suicidal Posts & Non-Suicidal Posts

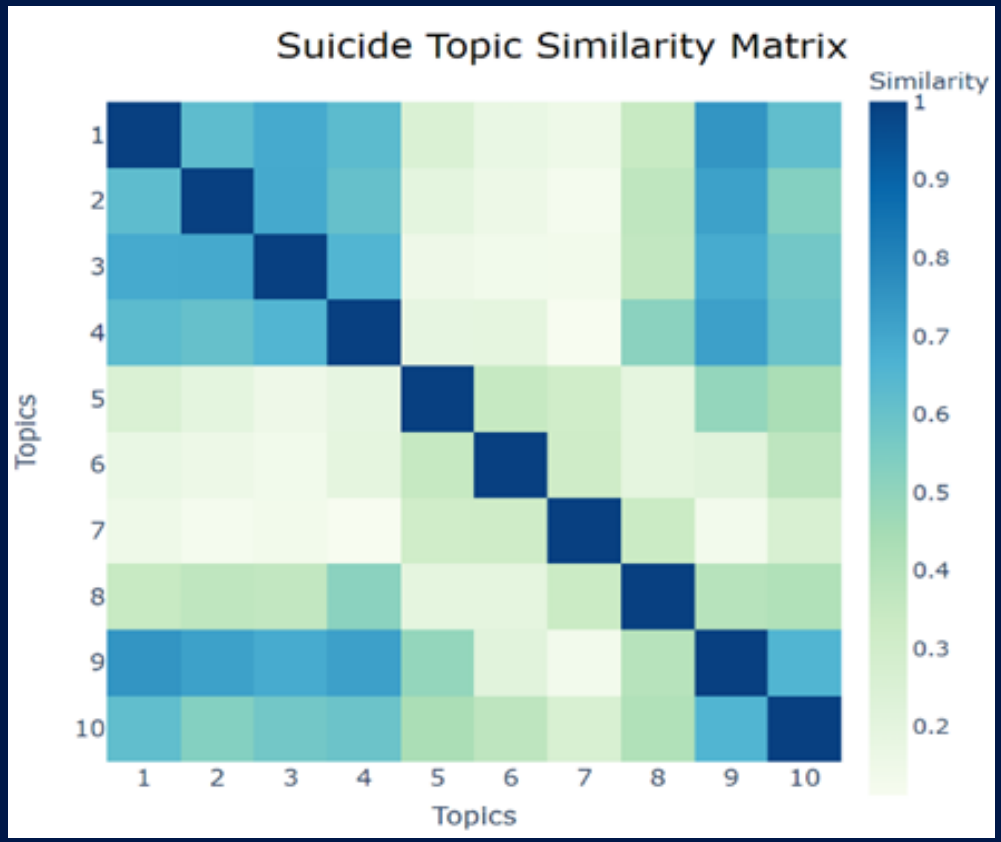
Rank	Topics		Topics	
	Suicide Topic (n=39742)	Count	Non-suicide Topic (n=63885)	Count
1	Struggles with breakups and emotional turmoil in relationships	3285	Navigating romantic relationships and feelings	6711
2	Struggles with academic failure and mental health	1791	Concerns with boredom and loneliness	3286
3	Childhood and family trauma	1421	Academic struggles for college students	1813
4	Emotional dilemmas with suicide	1273	Struggles with parents and family dynamics	1121
5	Suicidal intent with overdosing on pills	878	Sexual frustrations	611
6	Struggles with loneliness and low self-esteem	856	Challenges in socializing for introverts	536
7	Suicidal farewell messages	787	Insomnia and sleep struggles	440
8	Self-harm and suicidal ideation	538	Concerns over racism and racist language	431
9	Depressed birthdays and suicidal thoughts	392	Concerns with alcohol consumption	407
10	Suicidal intents with firearms	347	Struggles of transgender identity	393

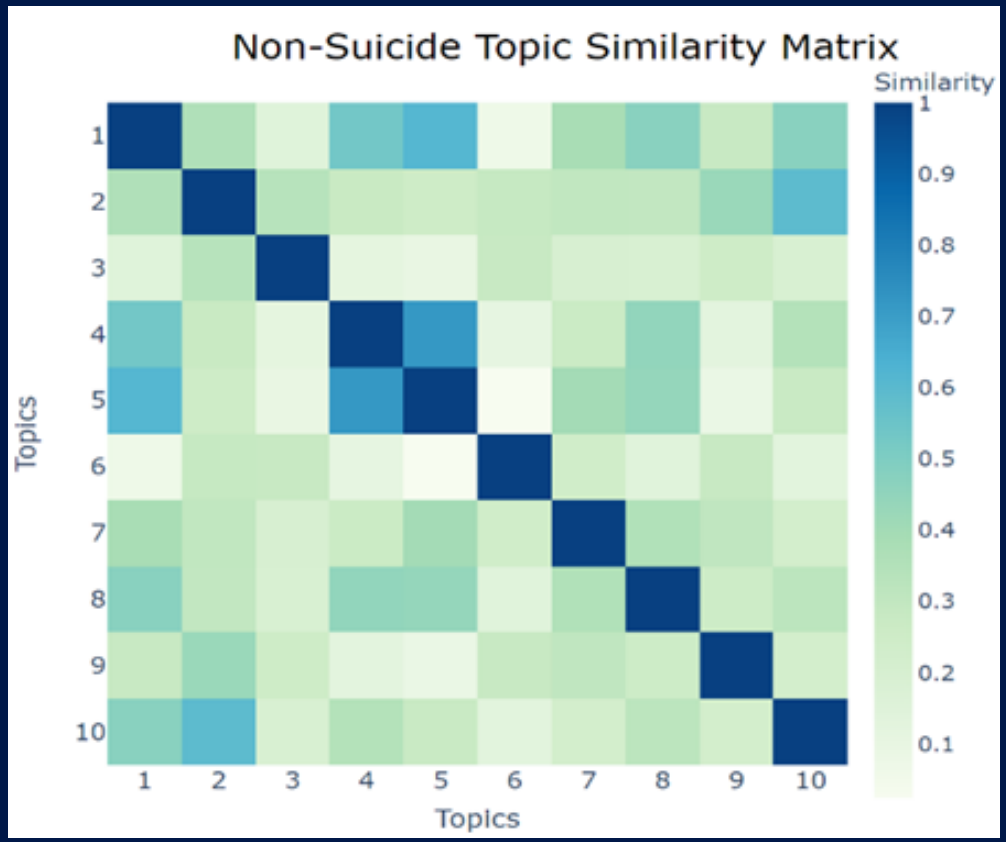


Results-Key Topics



Results-Key Topics





Conclusion

This study demonstrates the effectiveness of advanced NLP and statistical techniques in identifying and analyzing suicidal ideation based on large-scale social media data from Reddit. By fine-tuning Llama 3-8B, Mistral 7B, and BERTopic, we revealed key sources of mental instability associated with suicidal thoughts, including relationship struggles, academic stress, and family trauma. The findings also highlighted distinct thematic differences between posts indicating suicidal ideation and general adolescent concerns, revealing deeper insights into specific triggers and expressions of suicidal thoughts among young individuals. Our results underscore the potential of NLP in real-time mental health monitoring and intervention on social media. Fine-tuned models, such as Llama 3-8B, which achieved a test accuracy of 0.9371, demonstrate strong predictive performance, offering scalable tools for distress detection. By addressing warning signs before they escalate into crises, these systems can provide early interventions in teenage communities, leveraging the thematic overlaps between suicidal and non-suicidal groups to design broader mental health support initiatives. To enhance model accuracy, future research could incorporate data from a wider range of social media platforms to improve generalizability while also exploring the influence of digital interactions on users' mental health. Expanding this work will offer deeper insights into the evolving role of social media in mental health and help develop targeted intervention strategies tailored to specific online behaviors and mental health challenges.

Thank You!