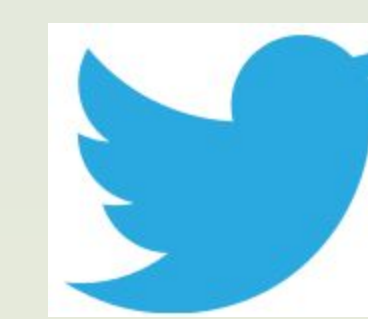


Suicidal Ideation: Social Media Detection



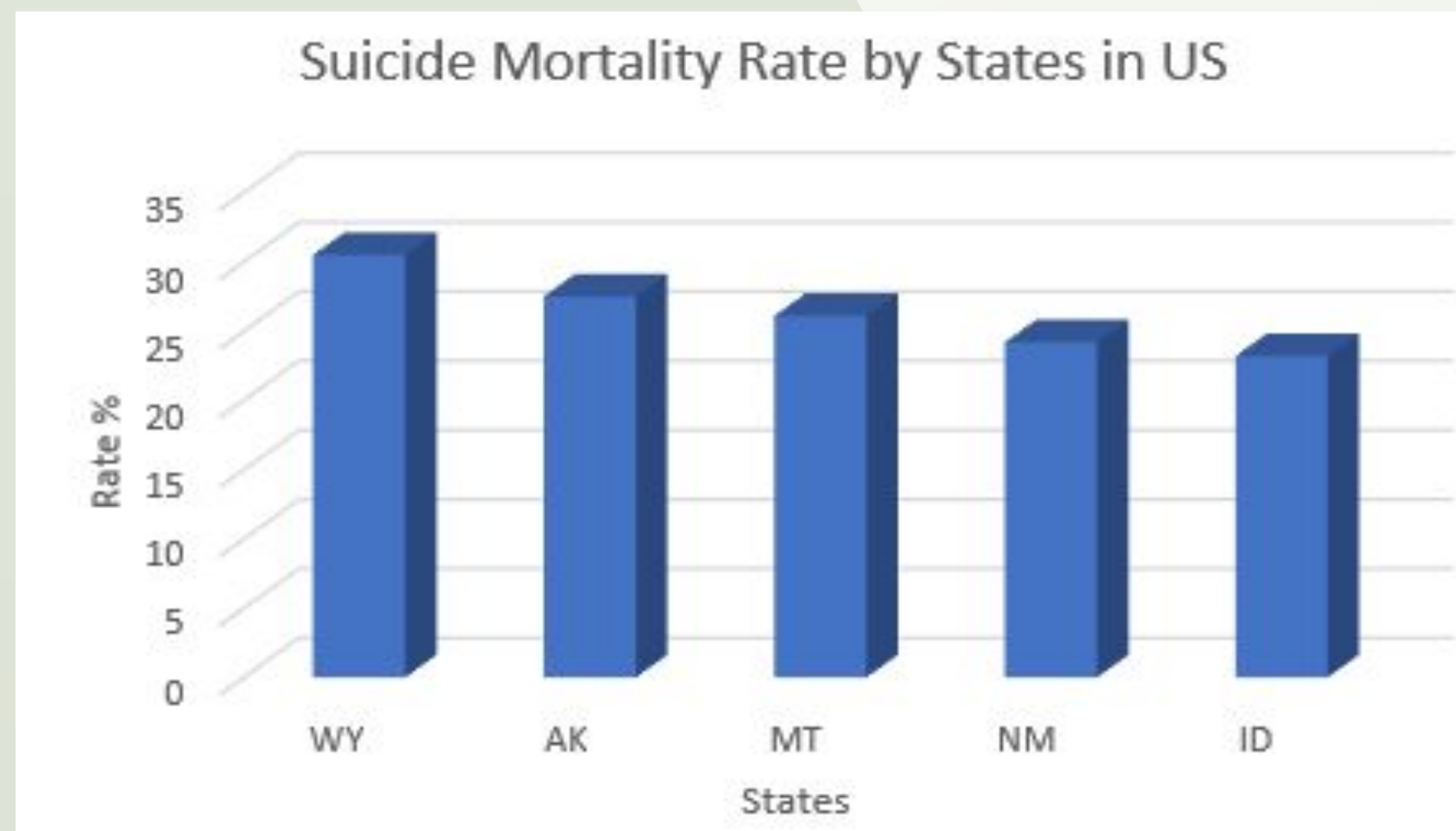
Team 32: Jessica Addai, Oluwabunmi Ariyo, Xiaoying Yang, Yue Huang, Countess Olufunmi Aforijiku, Helen Kpoto

TA: Shermeen Velani

Background

According to WHO, over 700,000 people take their lives each year and unfortunately even more individuals attempt suicide. Diagnoses and/or treatments are often delayed, not accurate or missed entirely.

In 2020, Wyoming had the highest suicide mortality rate in the United States.



This project will focus on tweets from residents of Wyoming, US.

Why This Matters

Since individuals with suicidal ideation are likely to signal their intentions, there is a high probability that these individuals voice out their feelings or thoughts in social media.



Datasets

- Suicide Detection from Reddit

232k+ post from “SuicideWatch” and “depression” subreddits of the Reddit platform used to extract key words related to suicide.

<https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch>

- Twitter API

Tweets from people containing keywords indicative of suicidal thoughts or behaviour based on historical data.

Exploratory Data Analysis

Text Preprocessing

- Converted all text to lower case
- Removed all digits
- Removed punctuation marks and emojis
- Filtered only english contents
- Word lemmatization

Top words related to suicide (left) and non-suicide(right)



Posts with suicidal tendencies contain more words about life, time, opposition/rejection compared to non-suicide posts.

Modeling

Training Model

- Multinomial Naive Bayes
- Scores

Accuracy	89.53%
Recall	97.18%

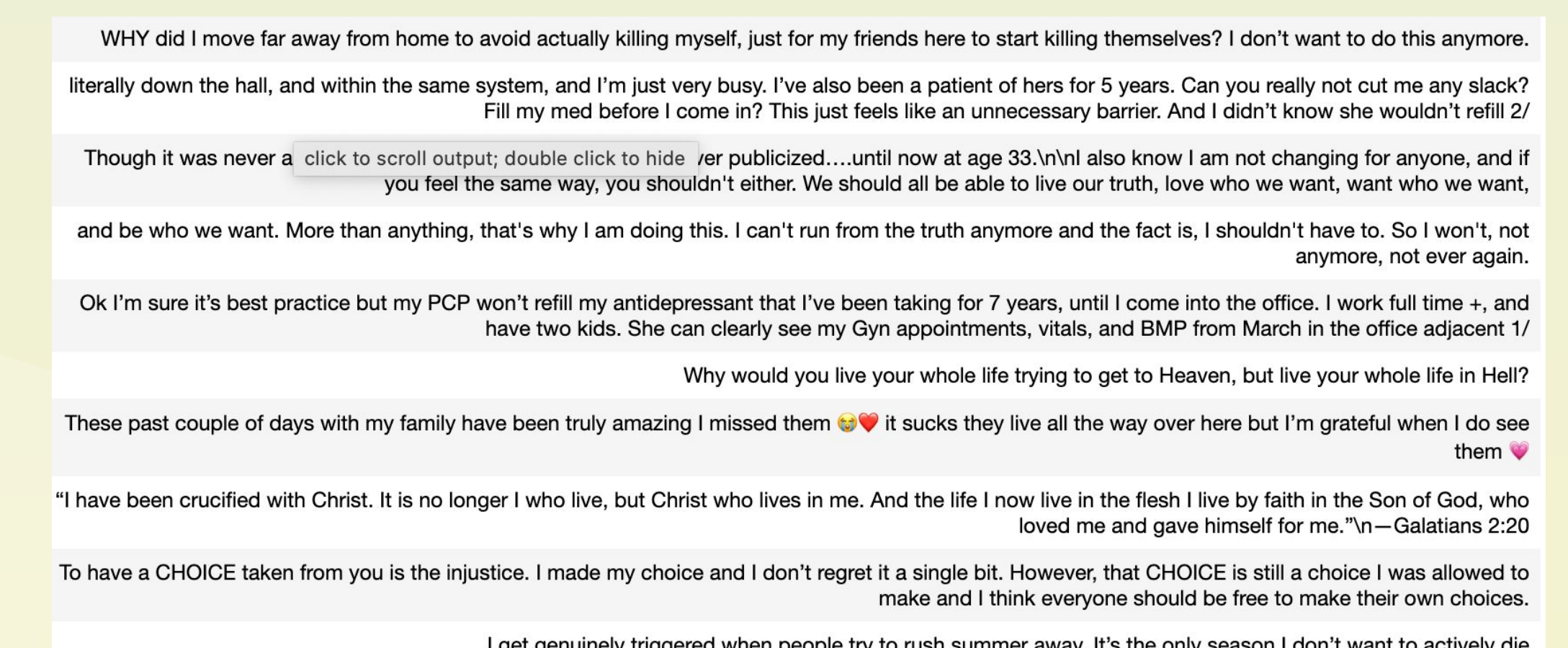
References

- <https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm>
- <https://www.ncbi.nlm.nih.gov/books/NBK565877/>
- <https://www.nimh.nih.gov/health/statistics/suicide>

Results

Model Deployment

- Tweet API (Wyoming data)
- The Top 10 Most Suicidal Tweets



- Suicide Risk Stratification Counting

Suicide Risk Stratification	Count
None	132
Low Risk	9
Moderate Risk	11
High Risk	15

Conclusion

- In 2020 ,amongst all states in the US, Wyoming had the highest suicide mortality rate at 30.5%
- Approximately 600 Wyoming Twitter data was scrapped for the month of June 2022. Based on the model used, 22.38% was classified as suicidal tweets.
- In assessing the average probability of suicidal tweets by different users, 15 users were flagged as high-risk, 9 users were flagged as moderate-risk and 11 users were flagged as low-risk of suicide tendency.

Next Steps

- Distinguish the multiple language characters in text data
- Work with emojis and special characters
- Extend to every other state In the US