Sentiment Analysis for Mental Health Issues

DSB 602 – Project 4

Anthony Chestnut, Beau Lambert, Drake Sigler

Mental Health By the Numbers

- 1 in 5 U.S. adults experience mental illness each year
- 1 in 20 U.S. adults experience serious mental illness each year
- 1 in 6 U.S. youth aged 6 – 17 experience a mental health disorder each year
- Mental health issues include:
  - Anxiety disorders
  - Depression
  - Post Traumatic stress
  - etc.

https://www.nami.org/about-mental-illness/mental-health-by-the-numbers/

The Ripple Effect of Mental Illness

- Increased risk of:
  - Cardiovascular and metabolic diseases
  - Substance abuse disorder
  - Suicide

- Families Affected:
  - Caregiver to adults with mental health issues (avg of 32hr/ wk.)

- Communities Impacted:
  - Increased homelessness
  - Increased hospitalization
  - U.S. Veterans and active service members suffer disproportionately
  - $193.2 billion in lost earnings each year

https://www.nami.org/about-mental-illness/mental-health-by-the-numbers/

Mental Health Resource Shortage

- Current state of mental health care in U.S.:
  - Mental health organizations are understaffed
  - People are unable to access or afford mental health services
  - Counselors are overwhelmed with high caseloads
- Reasons for mental health provider shortage:
  - Lack of funding → government provides a limited amount of funding
  - Poor reimbursement rates → providers are not adequately compensated
  - Low Retention → negative stigma associated with mental health care profession
  - Increased need, limited access → demand is outpacing the supply of providers
  - Aging workforce → retiring health care professions and low retention rate are outpacing younger professionals entering the field

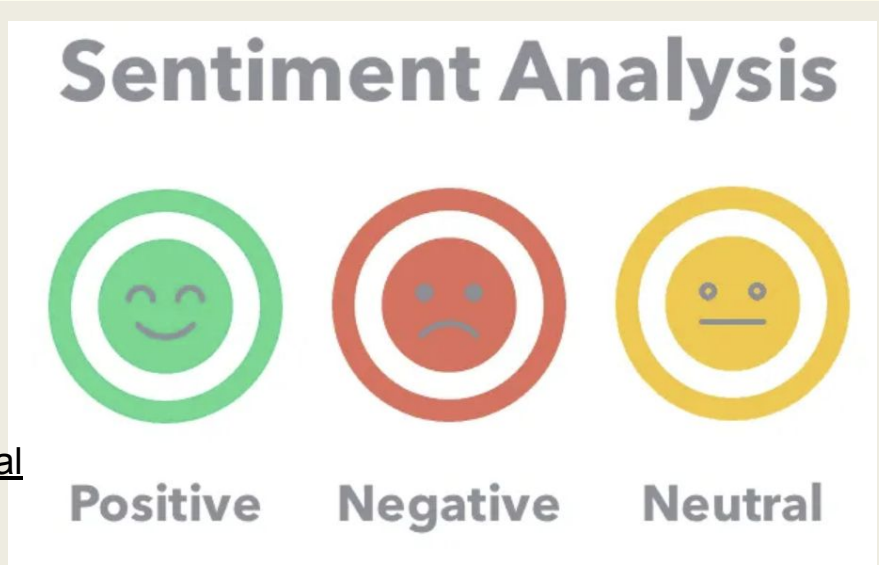
https://www.counseling.org/publications/counseling-today-magazine/article-archive/article/legacy/a-closer-look-at-the-mental-health-provider-shortage#

Problem Statement

- Millions of people in the U.S. are affected by mental illness each year
- However, resources for mental health limited in the U.S.
- Also, culture tends to suppress those who need help from seeking it
- It would be immensely valuable to provide a model which could predict whether an individual was suffering from various mental health issues
- Using sentiment analysis as a backbone for classification could be an effective indication of possible mental health concerns
- We'll use Recall/Sensitivity as our main metric since
    - False Positives would suggest people get help when they may not necessarily need (but everyone needs it) vs a false negative that could minimize serious health issues
- We'll set our target at least 95% recall before production-ready

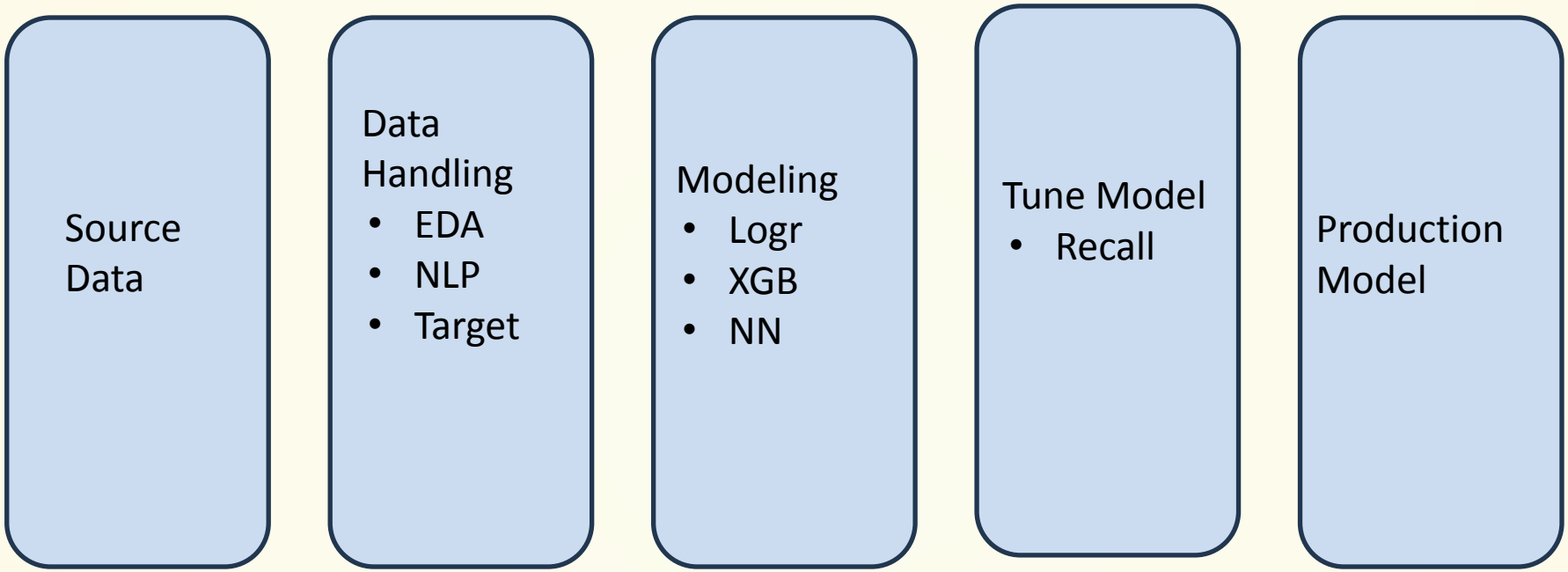
What is Sentiment Analysis?

- The process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral
- In today's world, there are enormous volumes of text data
  - Emails
  - Chat transcripts
  - Social media comments
- Sentiment analysis tools can scan this text to infer emotion
- Sentiment analysis can mine data information at great scale with reduced resources
- The limited resources can then be focused on the potential problem issues (negative cases)



https://aws.amazon.com/what-is/sentiment-analysis/
https://medium.com/@ahmettsdmr1312/customer-reviews-sentiment-analysis-two-different-techniques-21db5e67627b

Problem Solving Workflow



Sentiment Analysis for Mental Health – The Data

- The dataset integrates information from the following Kaggle datasets:
  - Depression Reddit Cleaned
  - Human Stress Prediction
  - Predicting Anxiety in Mental Health Data
  - Mental Health Dataset Bipolar
  - Reddit Mental Health Data
  - Students Anxiety and Depression Dataset
  - Suicidal Mental Health Dataset
  - Suicidal Tweet Detection Dataset
- Data Collection:
  - The data is sourced from diverse platforms including social media posts, Reddit posts, Twitter posts, and more
  - Each entry is tagged with a specific mental health status
- Features:
  - unique_id: A unique identifier for each entry.
  - Statement: The textual data or post.
  - Mental Health Status: The tagged mental health status of the statement.

https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data

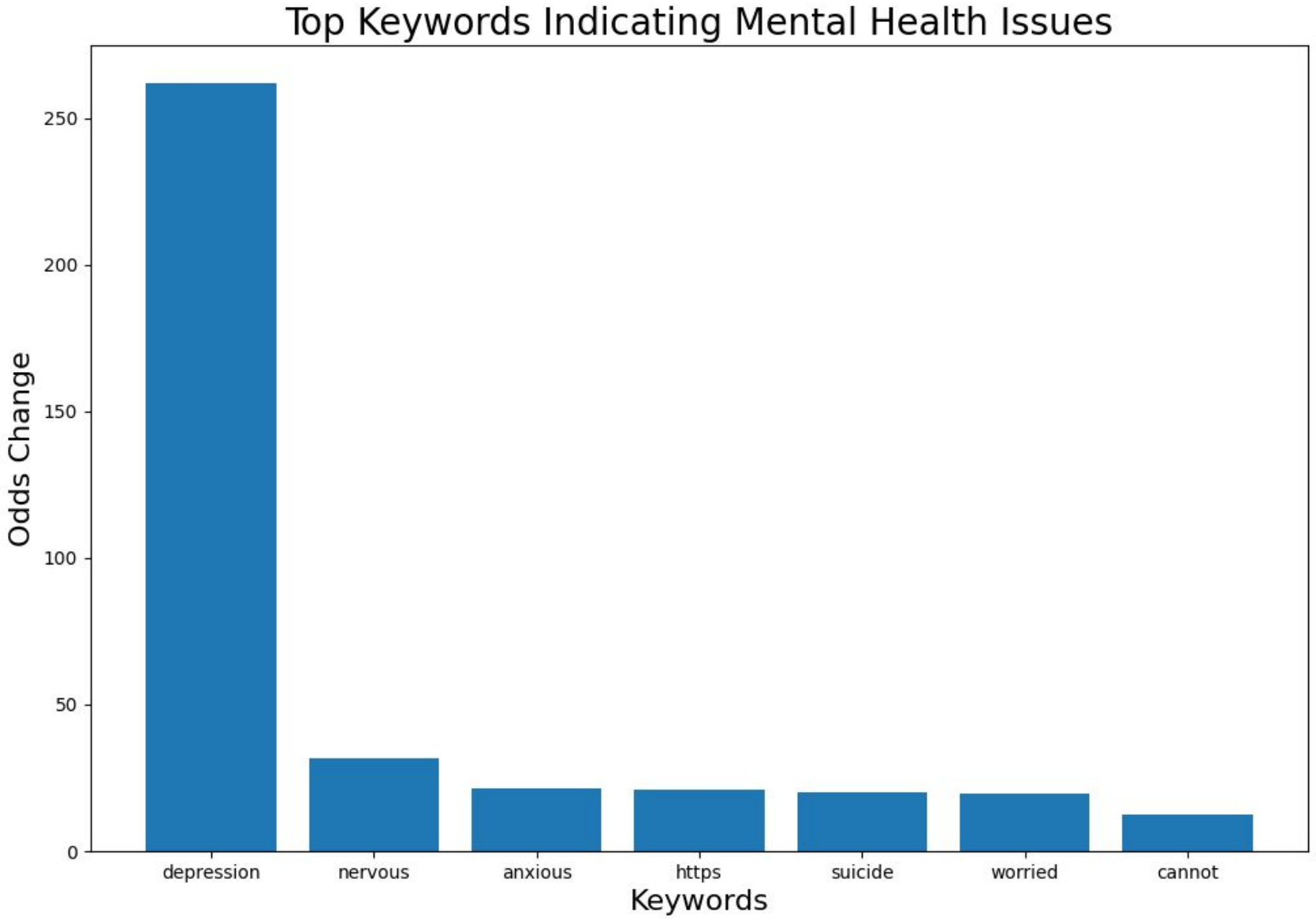
Data Handling

- Is the data usable?
  - Significant amount of missing or incomplete data
- How do we handle text data?
  - Pre-processing text → NLP (Natural Language Processing)
  - Breaking the text down into a form that can be put into a model
- How do treat the target?
  - We need to identify people that need help
  - Two classes
    - Normal
    - Distressed (possible mental health issues)

Modeling

- Technologies used
  - pandas
  - sklearn
    - CountVectorizer, TfidfVectorizer
    - LogisticRegression
    - MultinomialNB
  - Xgboost
    - XGBClassifer
  - Torch
  - **Bert**
    - **BertForSequenceClassification**
    - **BertTokenizer**
    - **Recall → 0.983**

Conclusions



Conclusions

- Data set was sourced that provided usable text data to be correlated with normal mental health or distressed mental health
- A model was developed that surpassed the metric of greater than 95% recall
- This proof on concept indicates that it could be possible to use sentiment analysis as a front-line indicator of possible mental distress cases that the current limited mental health resources could be focused on

Future Work

- Expand dataset with multilingual or real-time data
- Expand prediction capability with more than just binary classification
- Collaborate with mental health professionals for validation
- Build a dashboard or visualization interface for non-technical users

Questions?