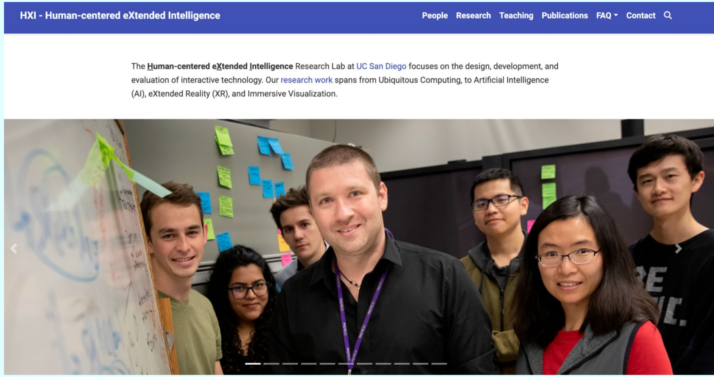
Predicting Reddit Post Topics for Mental Health Resources

Project Group #: 04
Nathaniel Do and Jesse Parent

Inception + Broader Background



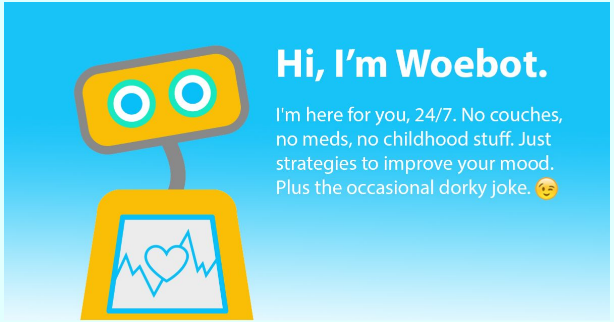
- A Different Kind of Capstone
  - Exploratory & preliminary work in service of larger aims

- The “UNSAD” Project
  - Dr. Nadir Weibel and his team at UCSD’s Human-centered eXtended Intelligence (HXI) lab have been working on UNSAD.
  - They scraped the subreddit r/dbtselfhelp, a online forum where users can ask for help with mental health issues using dialectic behavioral therapy, for data as well as followed the subreddit’s FAQ to find mental health resources over different topics.
  - Their goal is to use a Retrieval-Augmented Generation to create a AI “bot” that will interface with those seeking mental health support.
    - AI would be able to read a post from the subreddit, determine the topic, and recommend resources.

Challenges and Opportunities of AI in MH Support

AI chatbots & LLMs for mental health support are gaining traction (e.g., Woebot, Wysa, ChatGPT), but they remain controversial due to:

- Hallucinations: AI may generate misleading or harmful advice.
- Lack of Personalization: Models lack deep contextual understanding, making responses generic or inappropriate.
- Ethical Risks: Privacy concerns, biases, over-reliance on AI “companionship”, including at detriment of actual human professional therapy.



Potential Benefits of AI for Mental Health:

- 24/7 Accessibility—available when human professionals are not.
- Reduced Stigma—some users feel more comfortable talking to AI.
- Scalability—AI can reach many people at once, addressing mental health service shortages.

Our Project’s Role in this landscape:

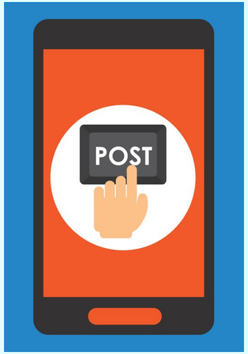
- To test how accurately ML models can classify mental health posts, laying potential groundwork for future RAG-based AI systems.

- Helps answer: *How well do current models recognize topic relevance?*

Challenges and Opportunities of AI in MH Support

- A more focused problem
  - People often post on social media platforms seeking help regarding mental health issues.
  - Responding to all of these posts would be too difficult for humans alone.

- Importance
  - After posting, many people may feel alone if they don’t receive the resources they need.
  - A snappy AI model can detect the topics of the posts to recommend targeted mental health resources.



Literature Survey & Use of ML For This Problem Space

Previous Work in Mental Health AI & NLP

- Supervised Learning:
  - Naive Bayes, SVM, Logistic Regression used for mental health text classification.
  - Limitation: Struggles with nuanced context & similar word usage across subforums.
- Unsupervised Learning:
  - LDA/NMF topic modeling explored to detect discussion themes in mental health forums.
  - Limitation: Topics often overlap, making fine-grained classification difficult.
- LLMs & Chatbots (Woebot, Wysa):
  - Strengths: Engaging, effective for basic cognitive behavioral therapy (CBT) responses.
  - Weaknesses: Risk of overgeneralization and AI-generated misinformation & Dependency/enmeshment

Our Contribution

- Instead of generating responses, **we analyze existing conversations** to study how mental health topics cluster in natural discussions.
- Establishes data-driven constraints that could help future AI systems **provide better, more relevant responses.**

Image: “Wysa: Everyday Mental Health”

Literature Survey & Use of ML For This Problem Space

💬 *When it comes to classifying and analyzing MH discussions,* ***rule-based keyword filtering isn't enough****. Traditional methods fail to capture* ***semantic relationships and emotional context****, which is why ML may offer a better solution.*

💬 *For example, simple keyword searches might classify a post with the word 'anxious' as belonging to r/anxiety, even though it may actually be discussing* ***relationship anxiety or PTSD symptoms****.* ***ML models learn from contextual cues*** *and can better distinguish between* ***similar but distinct topics****.*

💬 *Additionally, ML allows us to scale our analysis across* ***large amounts of Reddit data****, helping us understand the* ***structure of mental health discussions on a broader level****. This is particularly important if we aim to support* ***AI-generated responses in future MH chatbots****.*

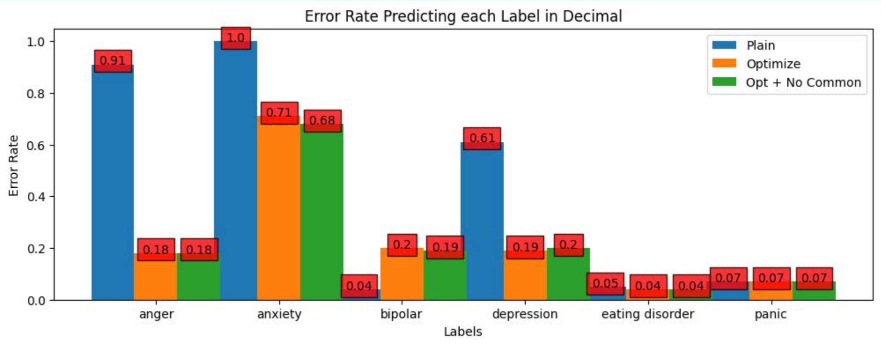


Image from our Results (Anxiety is Tough to Sort)

💬 *However, we also have to be mindful of* ***bias in AI models****. If our training data is skewed towards Western mental health frameworks, our AI might* ***misinterpret culturally specific discussions****. That's why part of our project involves* ***evaluating bias and refining our dataset to ensure fair classification****.*

Details on the Dataset

- UNSAD
  - Dr. Weibel and the HXI lab allowed us to view their dataset: text scraped from the subreddit r/dbtselfhelp.
  - As this is still a project in progress, the dataset is too small to use in ML.
- Branching Out
  - Seeing the general topics within the dataset, we scraped different subreddits across each topic to fill a dataset.
    - The raw data contained 12,951 rows of data with columns of the post’s title, body text, score, and topic.
    - After preprocessing, the dataset contains 9,078 rows of data with additional columns including processed title, text, and a combination of both title and text.
  - View Notebook!!!

Details on the Dataset

- UNSAD
  - Dr. Weibel and the HXI lab allowed us to view their dataset: text scraped from the subreddit r/dbtselfhelp.
  - As this is still a project in progress, the dataset is too small to use in ML.
- Branching Out
  - Seeing the general topics within the dataset, we scraped different subreddits across each topic to fill a dataset.
    - The raw data contained 12,951 rows of data with columns of the post’s title, body text, score, and topic.
    - After preprocessing, the dataset contains 9,078 rows of data with additional columns including processed title, text, and a combination of both title and text.
  - View Notebook!!!

Feature Extraction Details

- Preprocessing
  - View Notebook!!!
  - Basic
    - Lowercase all words
    - Remove:
      - Null values and numbers
      - Punctuation, emojis, and special characters
  - Advanced
    - Tokenize
    - Remove stop words
    - Lemmatization
    - Remove empty Rows

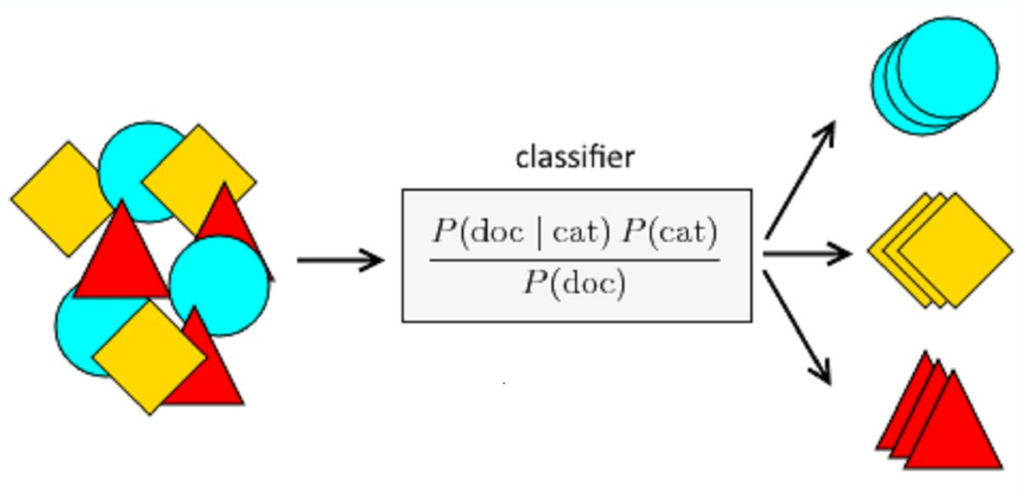
Feature Extraction Details

- EDA
  - View Notebook!!!
  - Common Words


```
anger   [('im', 1848), ('anger', 1457), ('like', 1377), ('get', 1202), ('feel', 1100), ('angry',
anxiety   [('anxiety', 981), ('im', 911), ('like', 711), ('feel', 667), ('get', 570), ('time', 4
bipolar   [('im', 3209), ('like', 2157), ('bipolar', 1844), ('feel', 1723), ('time', 1613), ('do
depression   [('im', 2988), ('like', 2024), ('feel', 2020), ('dont', 1812), ('life', 1274), ('wa
eating disorder   [('im', 2931), ('like', 2093), ('eating', 2034), ('dont', 1566), ('feel', 1536
panic   [('panic', 4844), ('attack', 4040), ('im', 2990), ('like', 2360), ('feel', 2178), ('anxi
```

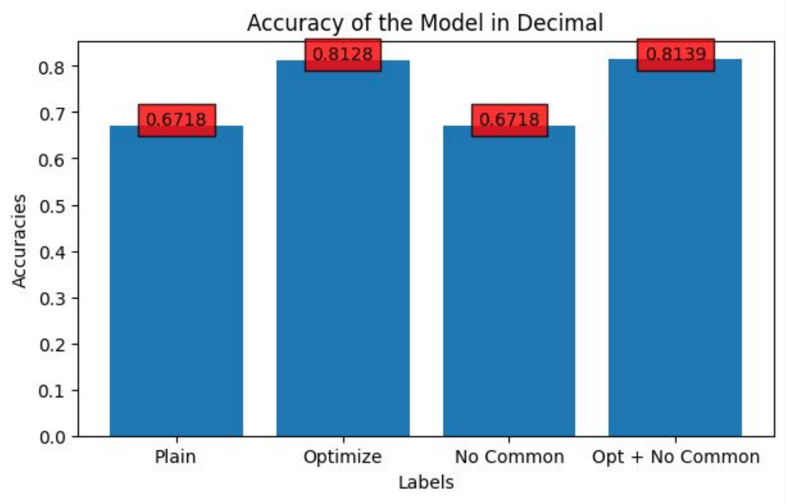
Model Details 1

- Multinomial Naive Bayes
  - Uses the frequency of a word appearing to predict the label.
  - View Notebook!!!



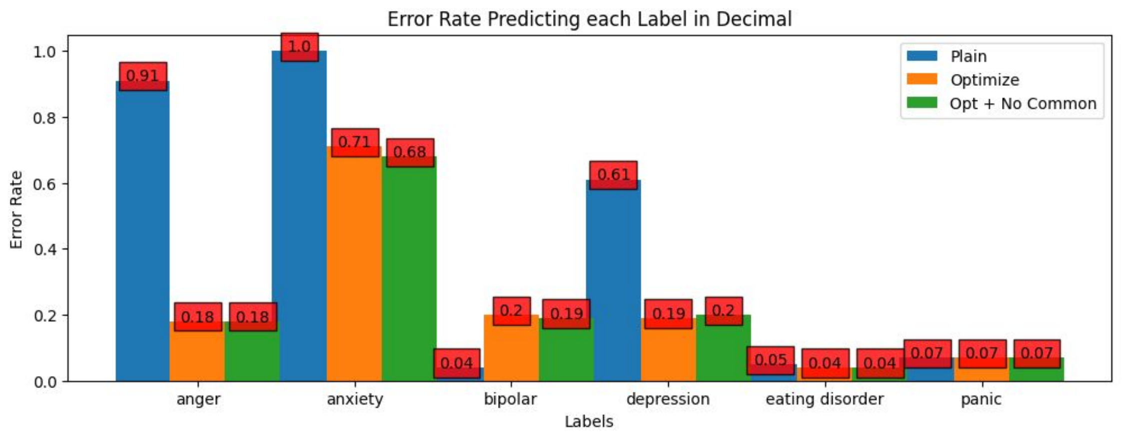
Results/Observations 1

- Multinomial Naive Bayes (Overall Accuracy)
  - Accuracy remains the same when removing common words.
    - 0.18% accuracy increase when removing common words from optimized model
  - 14.1% accuracy increase when optimizing parameters.
    - Parameter optimization is necessary while removing common words can be sacrificed if needed to speed up model.



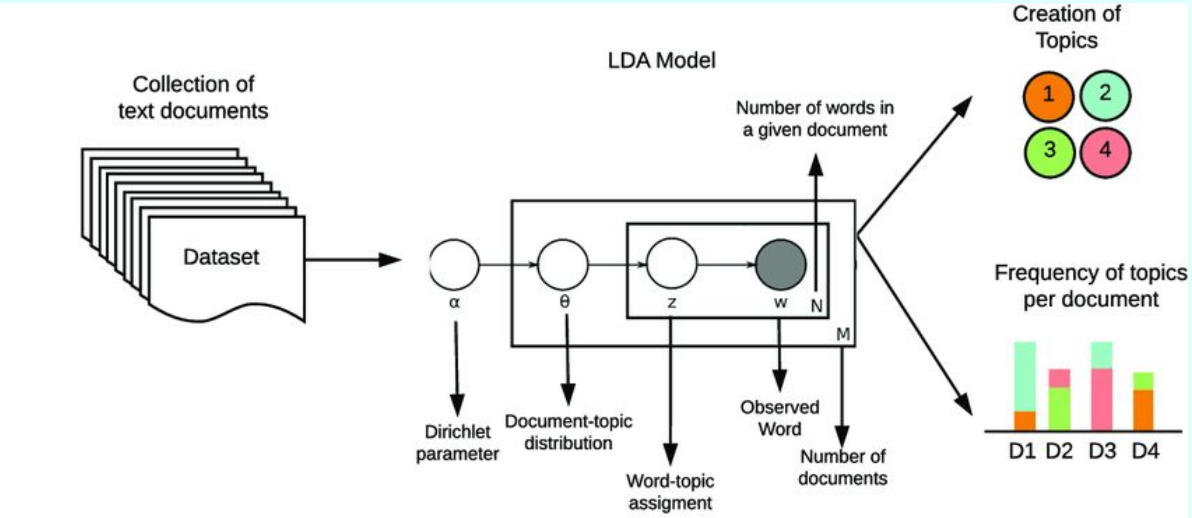
Results/Observations 1

- Multinomial Naive Bayes (Topic Error Rate)
    - Note: Left out “No Opt, No Common” model because of graph clutter.
    - Anxiety is almost always mislabeled and anger has a 91% error rate with the plain model.
    - Major decrease in error rate once optimized. Bipolar actually increases.
    - Small decrease in error rate once common words are removed except for depression.
    - Like the overall accuracy, optimization in necessary for a successful model while removing common words can be a negligible change.



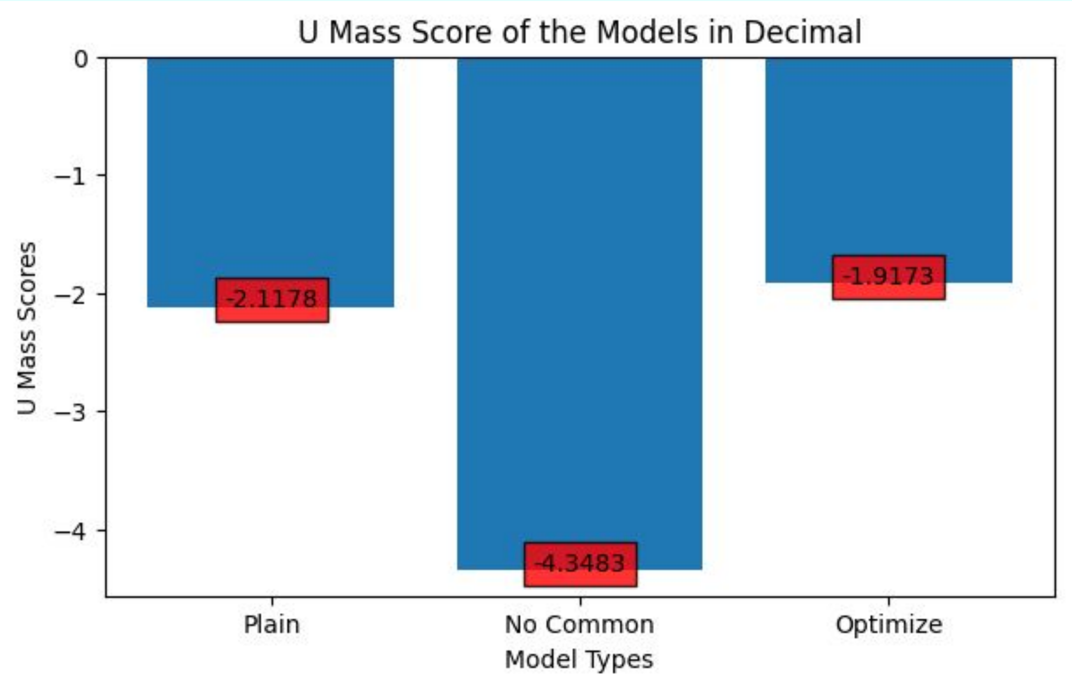
Model Details 2

- Latent Dirichlet Allocation
  - Give the model a number of topics to abstract to.
  - Scores the words based on how associated they are with the topics.
    - View Notebook!!! (Slow)
      - U Mass Score: Based on how often word pairs appear within a topic



Results/Observations 2

- Latent Dirichlet Allocation
  - Removing common words actually reduces the U Mass score



The Future

- Combination
  - With two solid models, we could search for a way to combine them to provide the most accurate topic prediction.
- Application
  - We have the makings of good Multinomial Naive Bayes and LDA models, but we have yet to test it on a subreddit not included in the training set.
    - We will apply our models to new subreddits to find how well our model does in the field.
    - Since posts may not be explicitly labeled “anger” or “depression,” we may have to read the field-test posts and assign labels ourselves.
      - Or we grab different kinds of posts from the same subreddit since we only grabbed the top posts.
- Resources
  - If time permits, we could gather various websites and resources to suggest once a topic is predicted.

The Future, Continued

- Discussion with HXI Lab about our findings
  - How can this help with the project as it stands?
  - What can we designate or document for them that will further future efforts?
  - RAGs vs other tools? What looks best now, a few months later?

- Documentation
  - Explanations and “Justification”, when possible

- Ethical Review and “Always-On Topics”
  - As from our course 266 Human-Centered AI, what topics
    - Privacy & disclosure/proper data use
    - Over-reliance on artificial companionship vs built-in efforts to individual, prevent enmeshment, and appropriate use relative to professional help

References & Thank You!

- Subreddits Used:
  - r/Anger, r/Anxiety, r/socialanxiety, r/bipolar, r/BipolarReddit, r/BipolarSOs, r/bipolar2, r/depression, r/depression_help, r/EatingDisorders, r/EDAnonymous, r/PanicAttack, r/panicdisorder
- Preprocessing Sites:
  - https://www.geeksforgeeks.org/text-preprocessing-for-nlp-tasks/#text-preprocessing-technique-in-nlp
  - https://codefinity.com/blog/A-Comprehensive-Guide-to-Text-Preprocessing-with-NLTK
- Modeling Sites
  - https://www.geeksforgeeks.org/multinomial-naive-bayes/
  - https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html
  - https://www.geeksforgeeks.org/latent-dirichlet-allocation/
  - https://www.geeksforgeeks.org/topic-modeling-using-latent-dirichlet-allocation-lda/