Predicting Depression in Reddit Posts with NLP

Spencer Hadel

Outline:

- ❑ Problem
- ❑ Data
- ❑ Methods
- ❑ Results
- ❑ Conclusion
- ❑ Appendix

The Problem

- In an age of increasing mental health awareness, we still struggle to accurately diagnose those in need, or even give them access to crucial care and support

- Because of lack of resources and information, people in need can go their entire lives without receiving the help they need

The Problem

Those in need of Mental Health care can be:

Undiagnosed

Even when they may be in dire need of treatment or a support network

Misdiagnosed

Due to lack of understanding of mental health, leading to more problems

Stigmatized

Due to a cultural misunderstanding of what a person needs to be safe and comfortable

The Solution

- **Reddit** will attempt to revolutionize mental health diagnosis, starting with its own platform and users

- Using posts from users with and without mental diagnoses, we can create a NLP model that can accurately identify traits of mental illness, and use the results to inform those in need

The Solution

- Before this system can be built, though, Reddit must first determine the viability of creating **accurate** and **ethical** models which can identify mental health outcomes
- Reddit can trust the results of this analysis to indicate that it is *worth the time and resources* to collect a massive, verified dataset of users' post history and mental diagnoses, in a variety of categories
- From there, we can revolutionize access to mental health diagnosis, options, and care

Data

- **7,731** reddit posts from users as **with** and **without** depression
- **3,831 ‘is_depression = 1’**

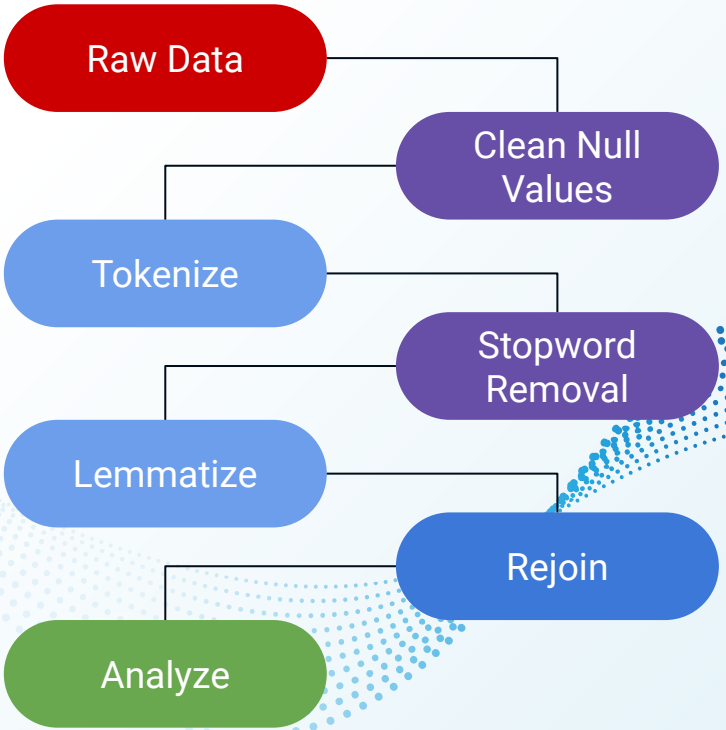
  3,900 ‘is_depression = 0’

NOTE:

Diagnosis is not verifiable based on this dataset!

Methods - Data Preparation

- **Tokenize:** separate words in the corpus into individual “tokens”
- **Stopwords:** “filler words” that don’t hold NLP value
- **Lemmatization:** reduce words to their “root”



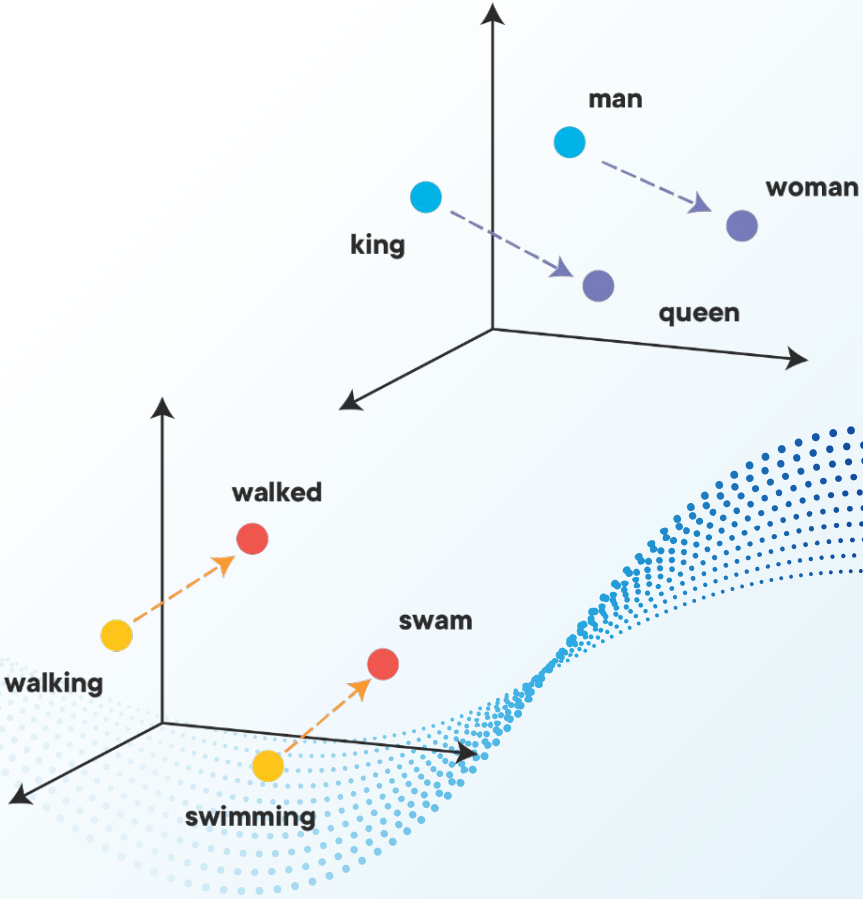
Methods - Modeling

- **Binary Classification Models**

  Identifying individual word weights

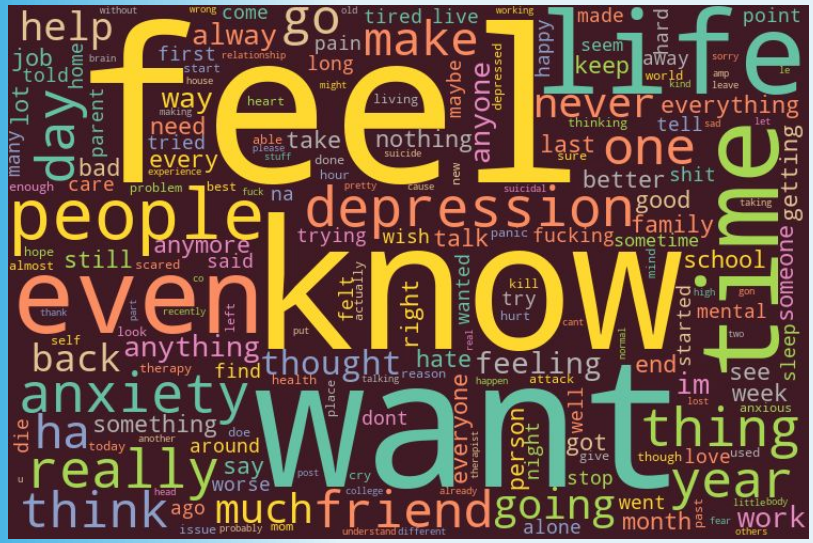
- **Word Vector Embeddings**

  Identifying words based on semantic relationships

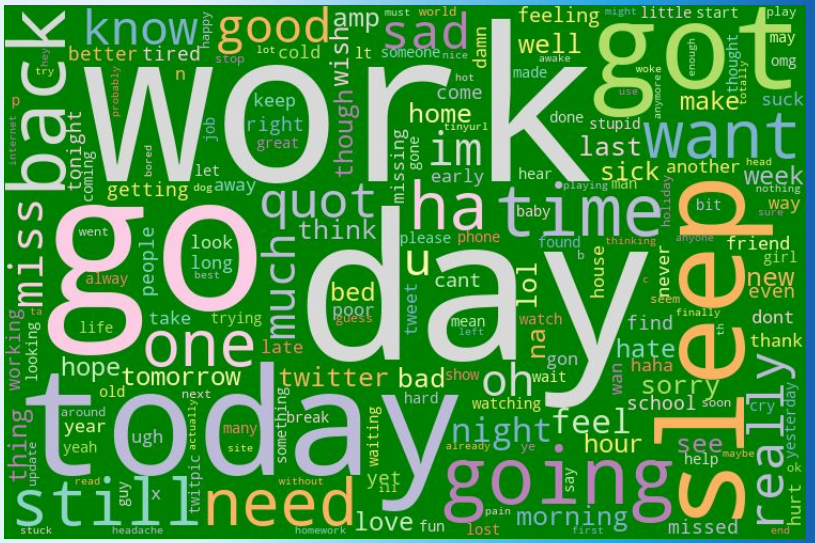


Results - Word Clouds

“Depressed” Class

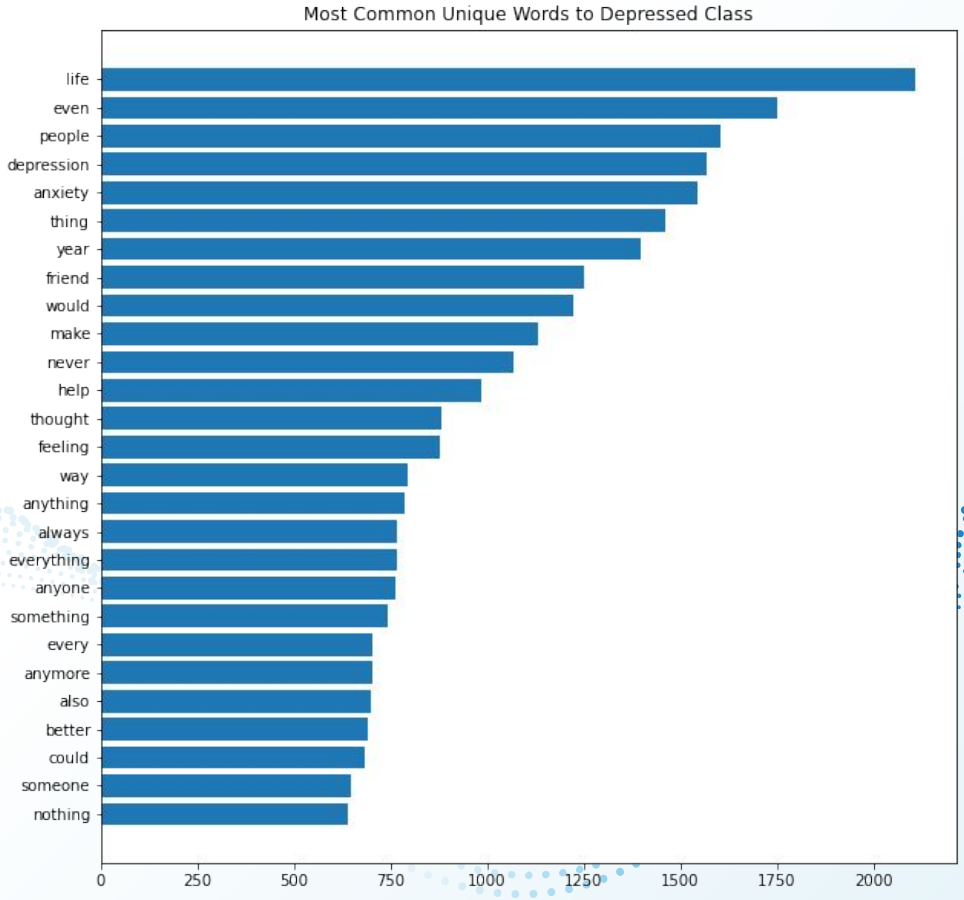


“Undiagnosed” Class



Results

- Initial analysis of the unique “depressed” data shows prominent word use



Results - Baseline Model

Accuracy:

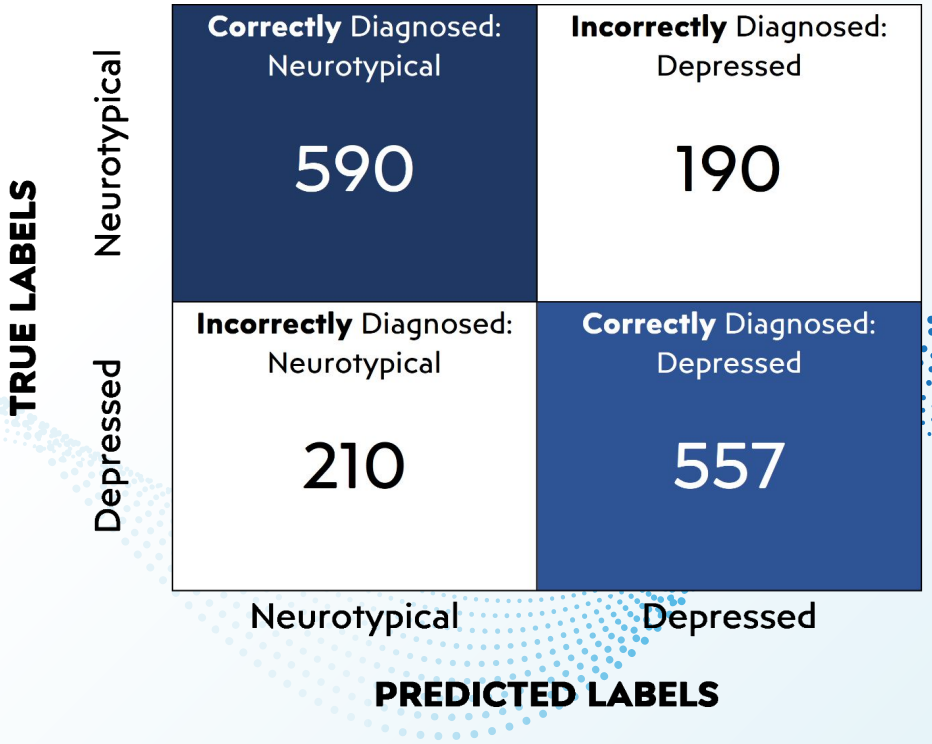
Percentage of *successful predictions* by the model:

74%

Sensitivity:

Percentage of true values (is_depression = 1) *successfully identified*.

72%



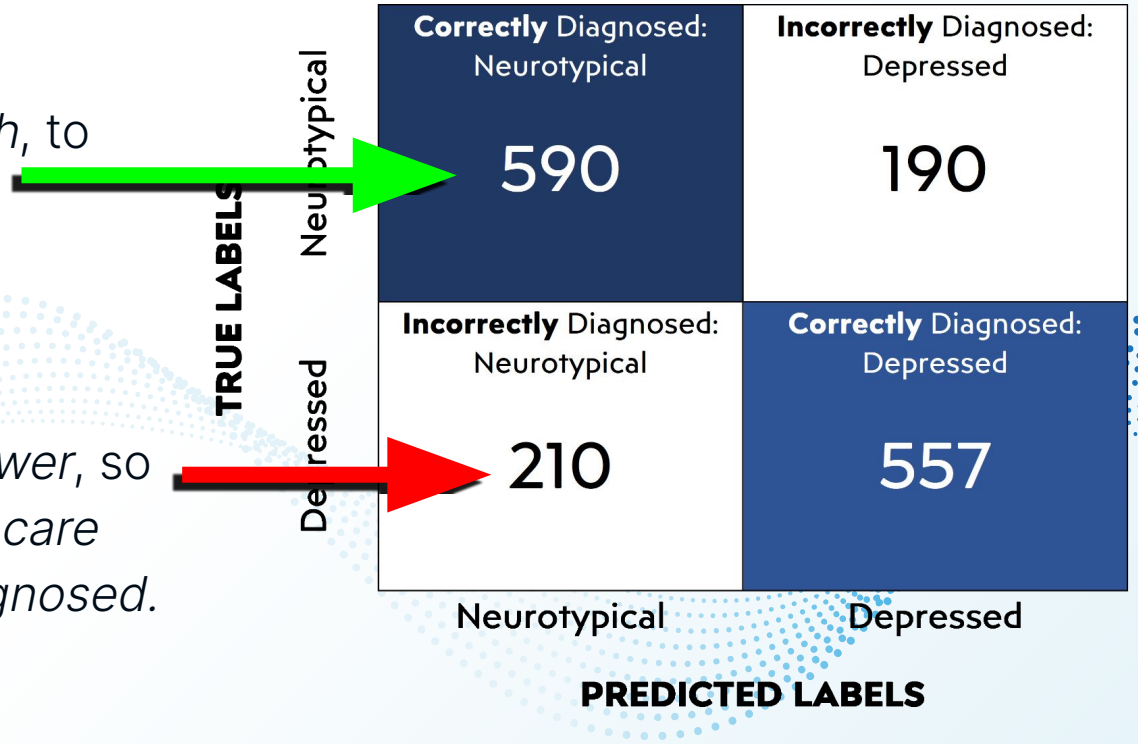
Results - Baseline Model

True Negatives:

Want to *keep this high*, to *avoid misdiagnosis*

False Negatives:

We want this to be *lower*, so that those *in need of care* are *appropriately diagnosed.*



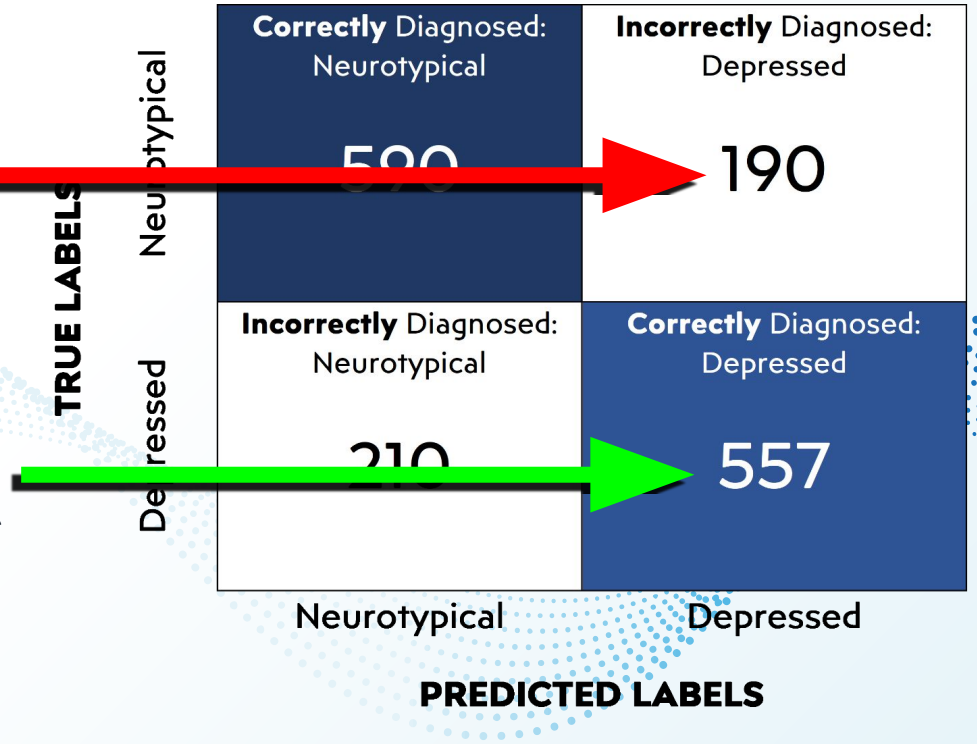
Results - Baseline Model

False Positives:

We also want *this to stay low, also to avoid misdiagnosis*

True Positives:

Higher is better, as each one indicates *a correct diagnosis of depression.*



Results - Final Classification Model

Accuracy:

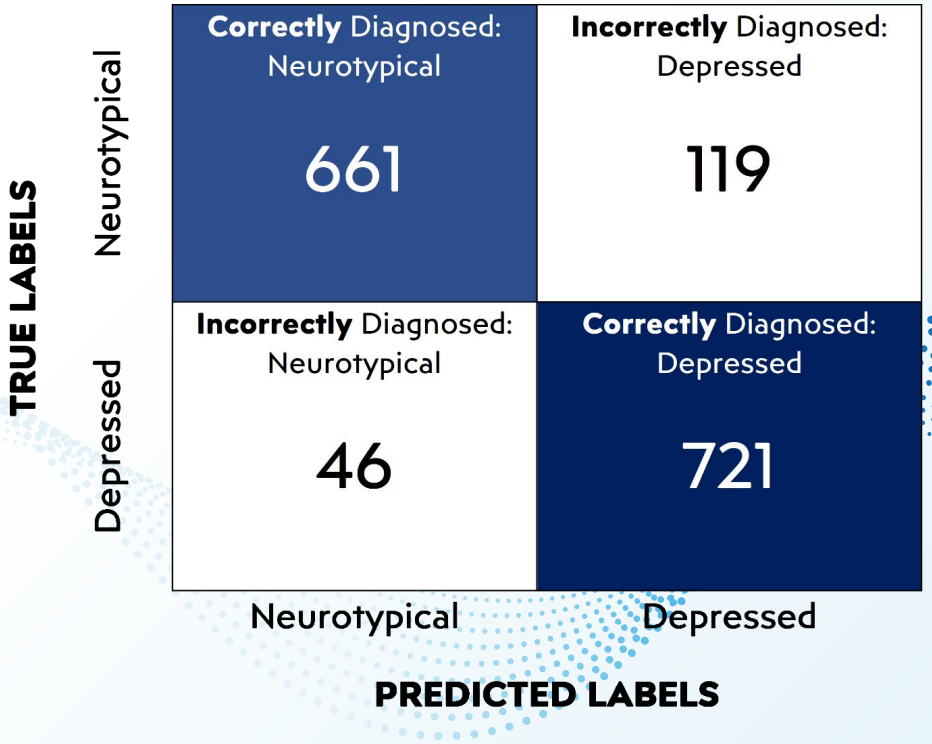
Percentage of *successful predictions* by the model:

89%

Sensitivity:

Percentage of true values (is_depression = 1) *successfully identified*.

94%



Results - Final Classification Model

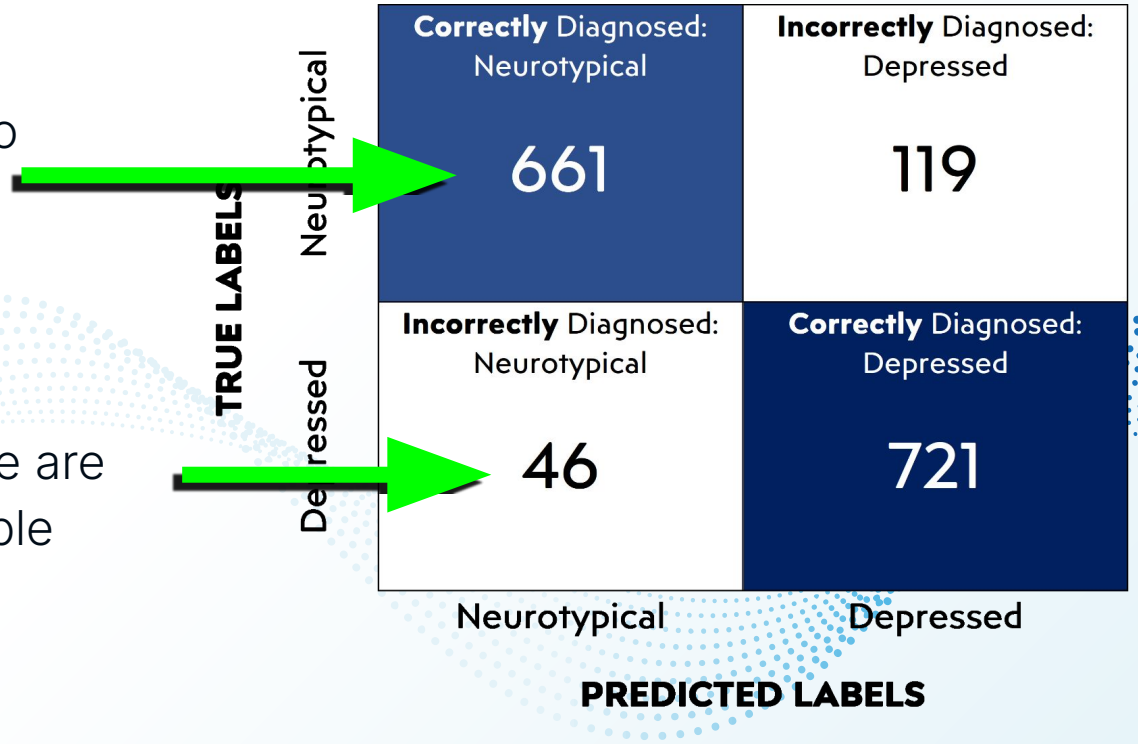
True Negatives:

Stayed high while also *drastically reducing*

False Negatives

False Negatives:

Improvement, as there are significantly less people going *undiagnosed* incorrectly.



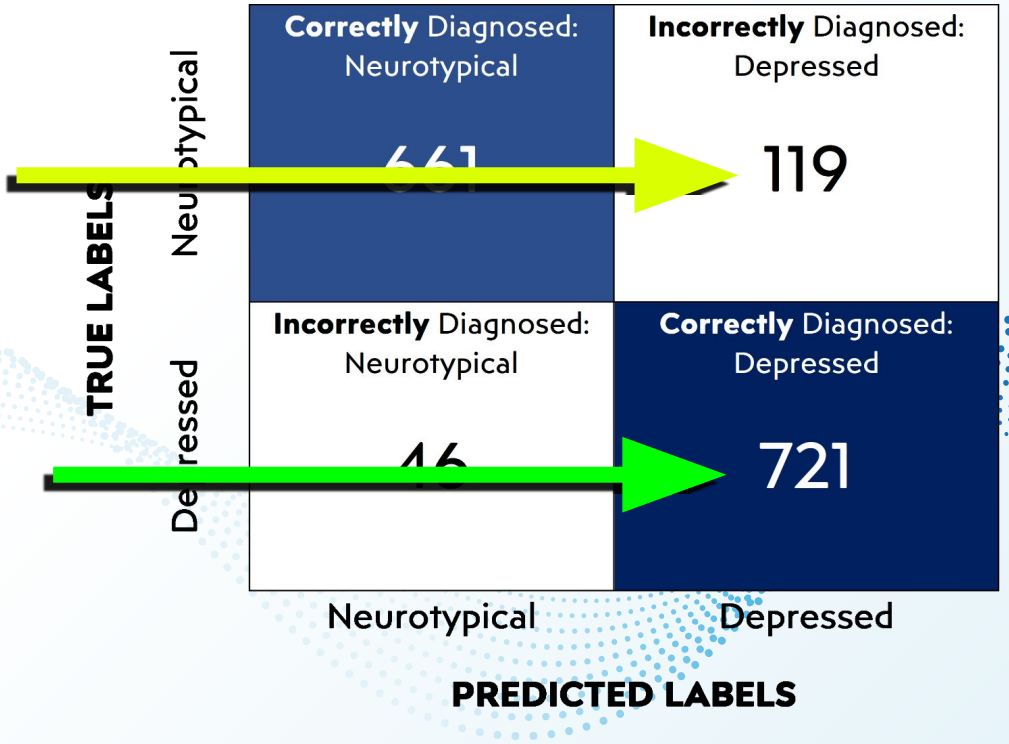
Results - Final Classification Model

False Positives:

Reduced by a bit but still *room for improvement*

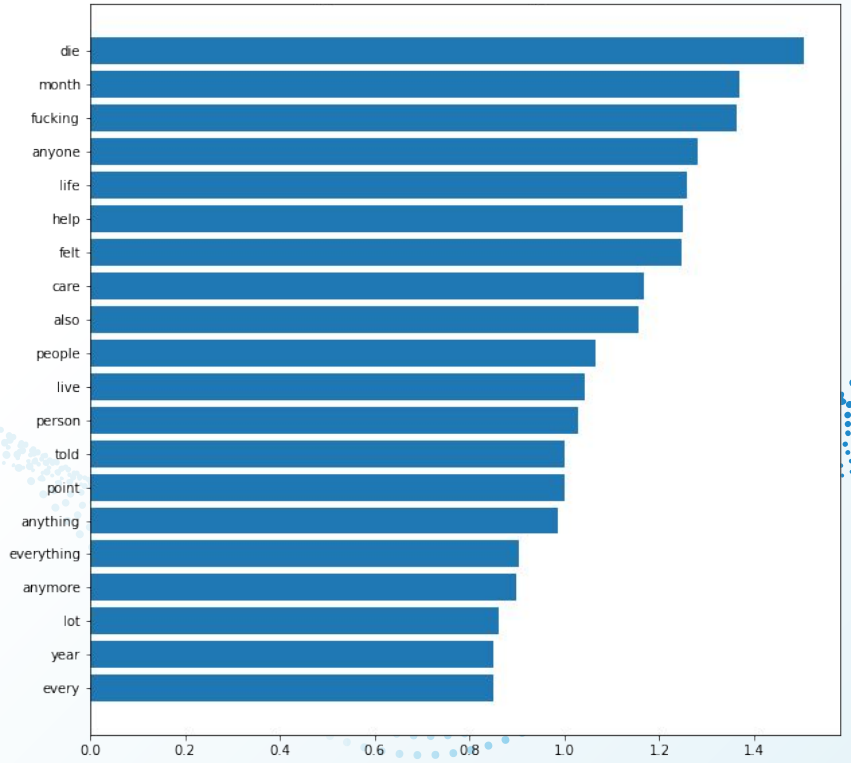
True Positives:

Final model is *far more accurate than baseline*



Results - Strongest Coefficients

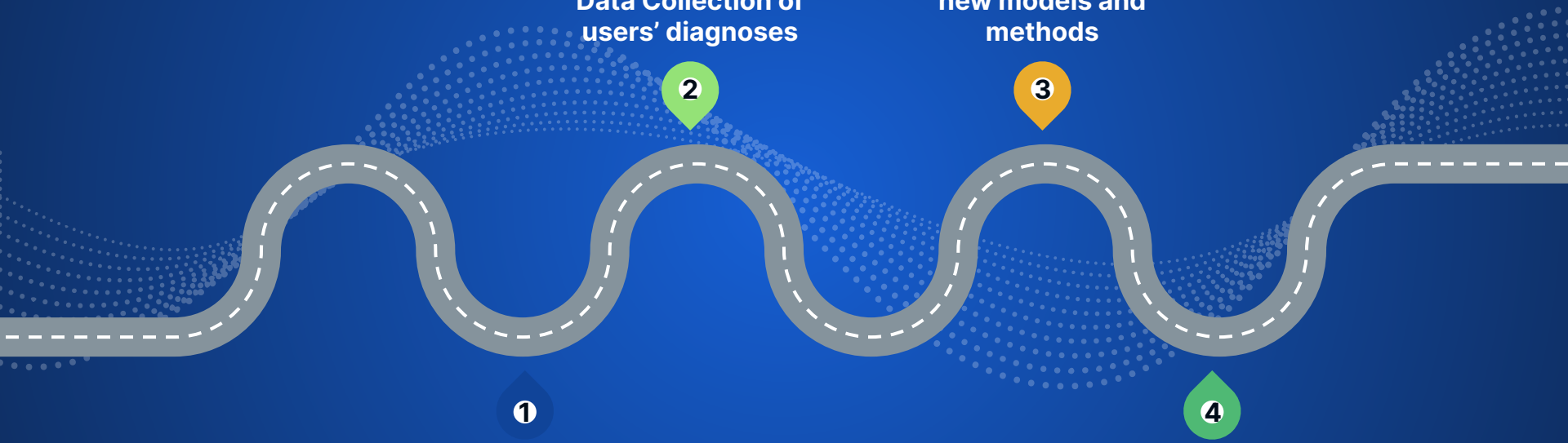
- The final model scored these features as the strongest coefficients, indicating that they had the greatest impact on determining “depression” in a post



Conclusion

- The results of this analysis clearly indicate that there is an identifiable trend in writing styles between those who are diagnosed with depression and those who are not.
- Reddit should *undoubtedly* go forward with large scale data collection in order to revolutionize public access to mental health care diagnosis.
  - For individuals, employers, teachers, parents, and mental health professionals around the world
  - Needs to be treated ethically and must avoid creating more stigma around mental illness

Next Steps



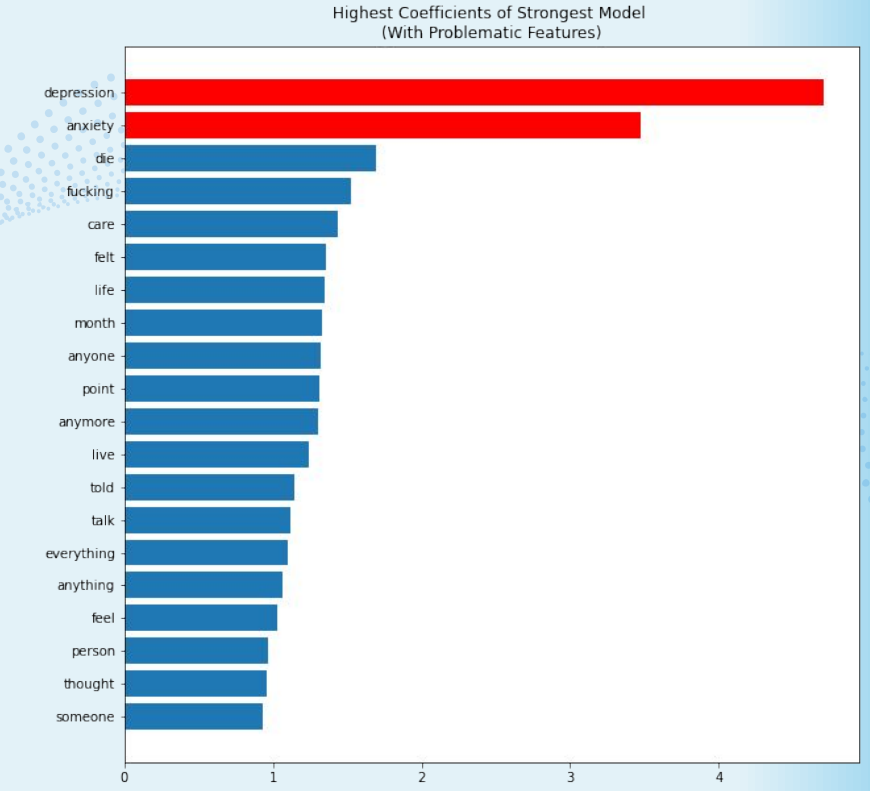
1. Apply current findings to a smaller scale program which can help people who are in need now
2. Begin Large Scale Data Collection of users' diagnoses
3. Continue testing current dataset with new models and methods
4. Run similar analyses on the few existing similar datasets (personality type datasets)

Thank You

Email: shadel96@gmail.com
LinkedIn: linkedin.com/in/spencer-hadel/

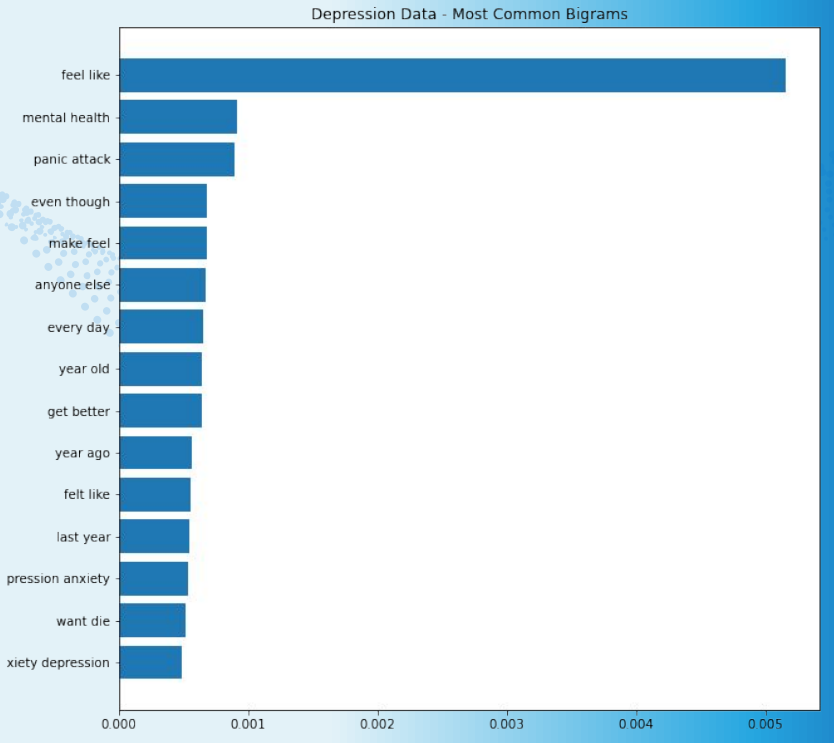
Full Analysis on Github:
https://github.com/shadel96/Mental_Health_NLP

Additional Results: Coefficients including Problematic Words



Additional Results: Bigrams in Depressed Class

- This visualization shows the most influential bigrams (word pairings) within the “depressed” class



Additional Results: Main Classifier Scores

Name	Accuracy	Recall	Precision	F1
SVC	0.893342	0.940026	0.858333	0.897324
Random Forest	0.878474	0.855280	0.894952	0.874667
LogReg	0.872010	0.915254	0.840719	0.876404
Tuned MNB - Acc	0.864900	0.963494	0.803261	0.876111
Decision Tree	0.824822	0.846154	0.809227	0.827279
Baseline Adjusted Data	0.741435	0.726206	0.745649	0.735799
Tuned MNB	0.735617	0.990874	0.654045	0.787973

Additional Results: Word Vectorization Model

Accuracy:

Percentage of *successful predictions* by the model:

92%

Sensitivity:

Percentage of true values (is_depression = 1) *successfully identified*.

87%