



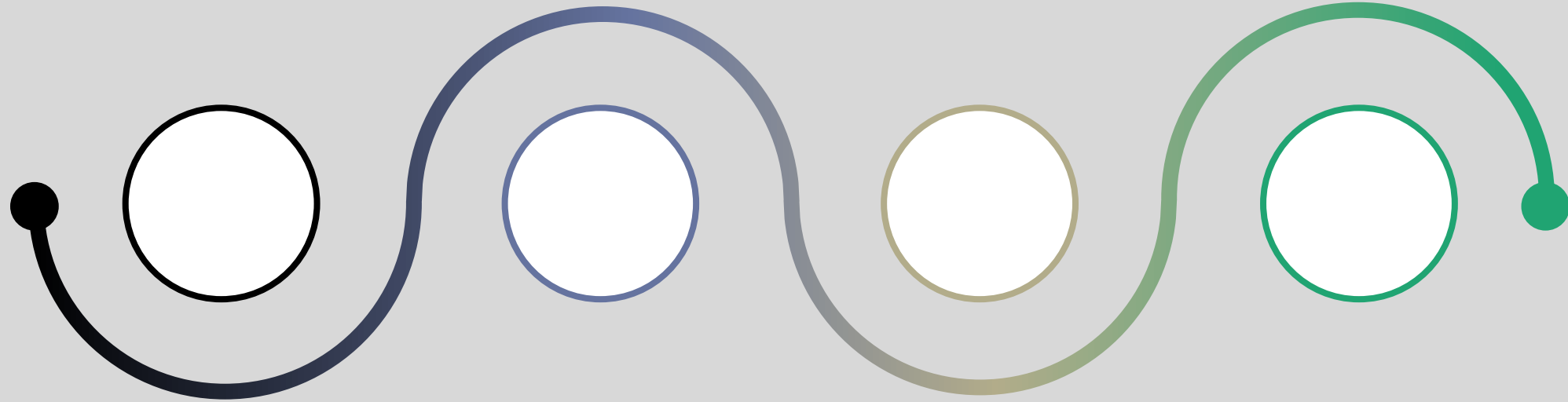
HELP SOCIAL CENTERS ACCURATELY IDENTIFY PATIENTS HAVING PTSD FROM ANXIETY

Audience: Non-technical social workers from counselling centers

Background: PTSD is a type of Anxiety problem, but not all with Anxiety has PTSD. Distinguishing factor is Trauma.

Aim: To provide model and incorporate into business process to accurately identify people with PTSD so that quick remedy can be provided.

Outline



Exploratory Data Analysis

Highlight key findings and trends

Classification Modelling

Classification Metrics

Models:

- Logistic Regression
- MultinomialNB
- Random Forest

Error Analysis

Key findings and improvement

Model Selection

Production model Selection

Key insights

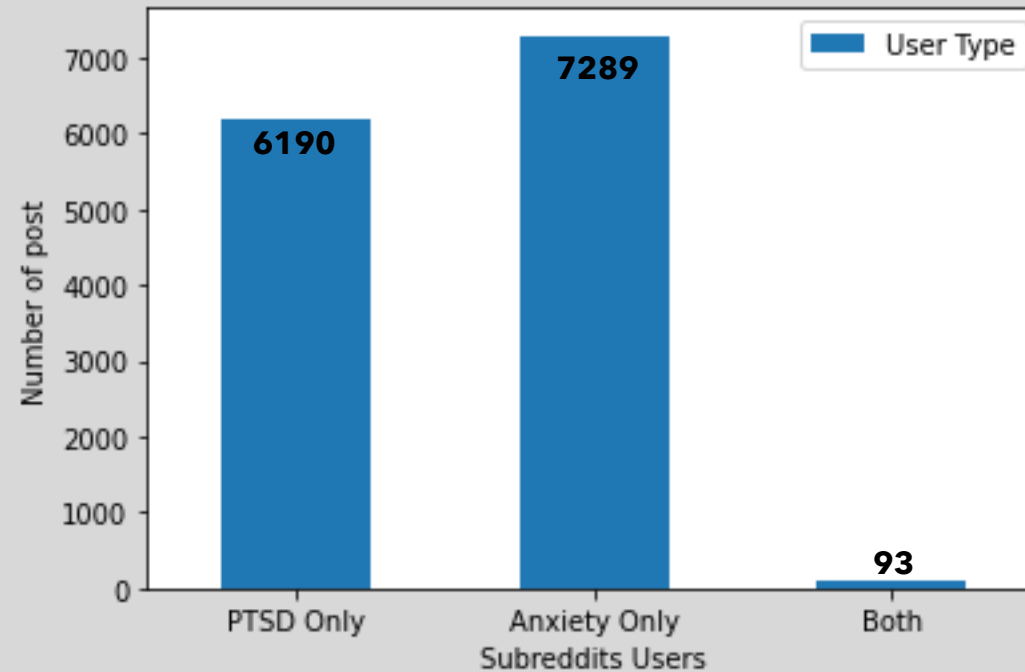
Exploratory Data Analysis

- Data imported from subreddits r/PTSD and r/Anxiety
- Pre-processing was done
 - Remove HTML
 - Remove all numbers and punctuations
 - Convert to lowercases
 - Removed Stopwords (common words)
 - Remove target leakage words such as trauma, PTSD, anxiety etc.

Visualizing differences in length of text

Visualizing Unique Users activity levels

Visualizing timeline across both subreddits VS No. of Posts

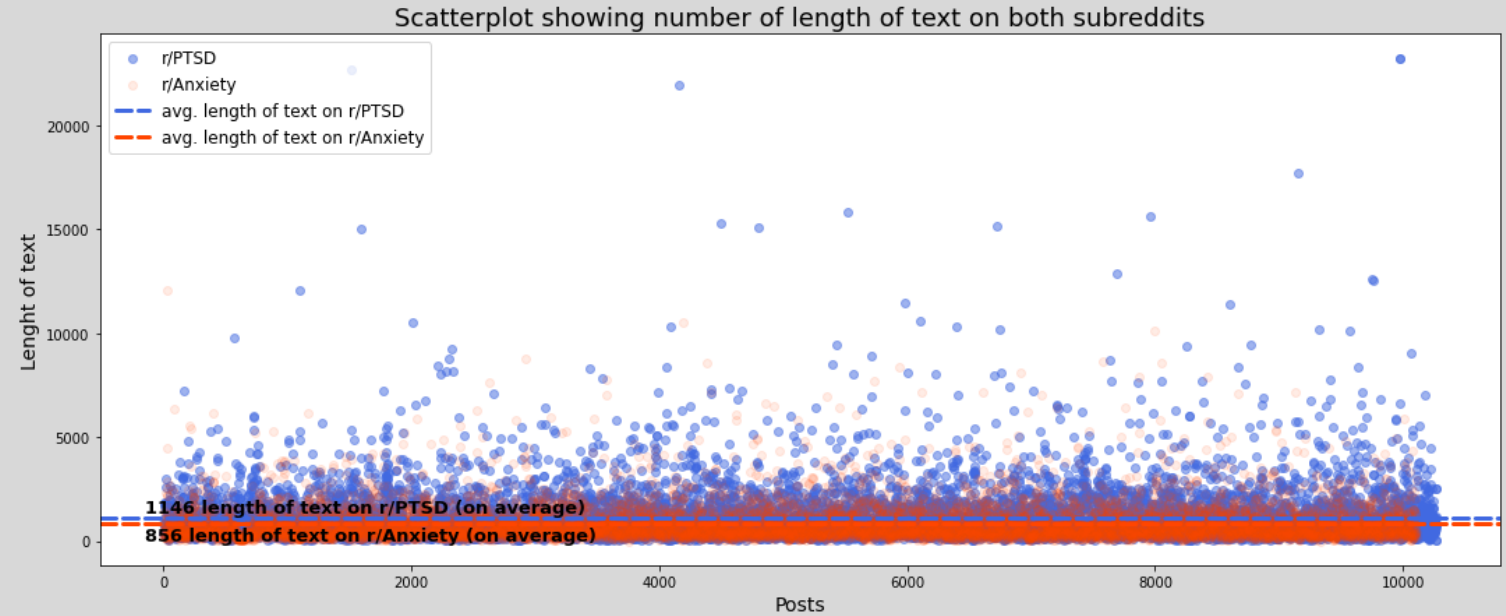


- Total of 93 users in both subreddits
 - These users are ambiguous with no clear indication whether they do have PTSD or Anxiety

Visualizing differences in length of text

Visualizing Unique Users activity levels

Visualizing timeline across both subreddits VS No. of Posts



- Length of text,
- Number of words
- Higher in r/PTSD than in r/Anxiety.
- Might indicate that people with PTSD have a lot more vent/experience/advice to face such issues

Visualizing differences in length
of text

Visualizing Users activity levels

Visualizing timeline across both
subreddits VS No. of Posts

➤ **Average no. of post per user**

- r/PTSD: 5
- r/Anxiety: 7

➤ **Average no. of post per hour**

- r/PTSD: 1.23
- r/Anxiety: 3.03

Generally..

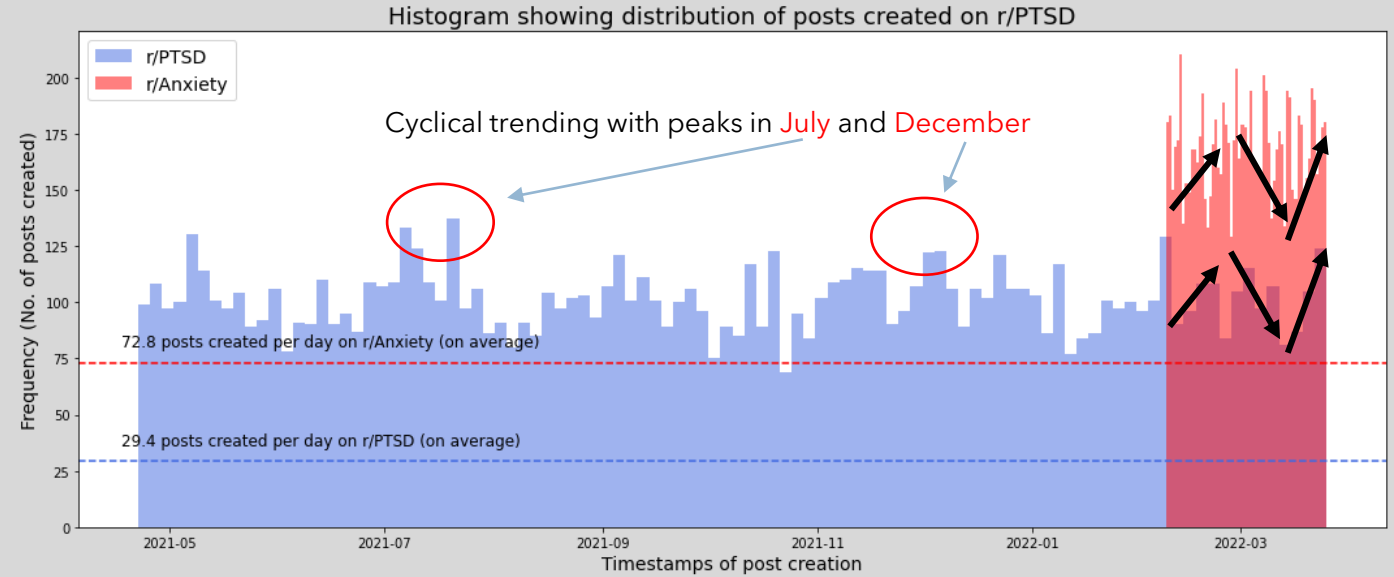
We can infer that:

- is likely more chance to find someone with Anxiety than having PTSD ([source](#))
- Users from r/Anxiety are more active by 2.5 times.

Visualizing differences in length
of text

Visualizing Unique Users
activity levels

Visualizing timeline across both subreddits VS No. of Posts

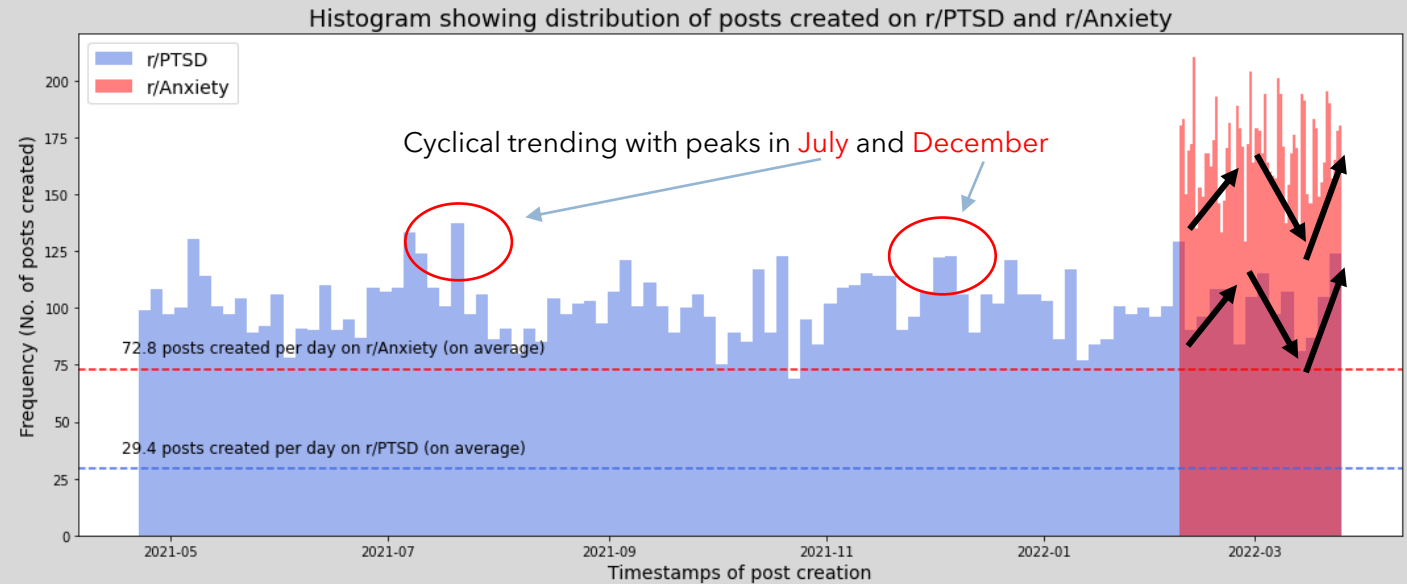


- Cyclical Trend for number of post across months
- Observed peaks in July and December

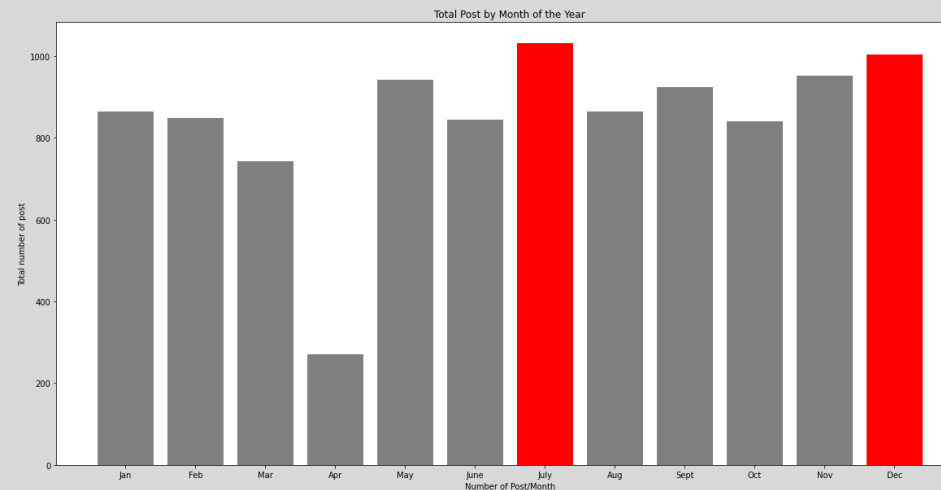
Visualizing differences in length of text

Visualizing Unique Users activity levels

Visualizing timeline across both subreddits VS No. of Posts



- Cyclical Trend for number of post across months
- Observed peaks in July and December

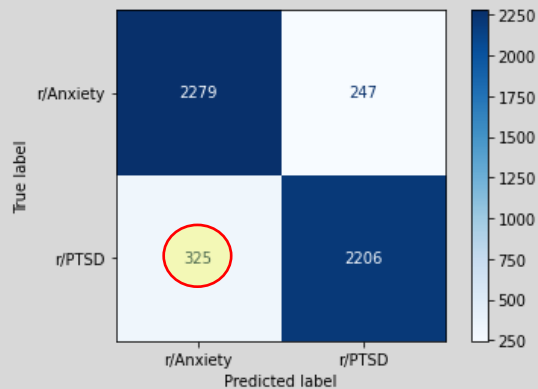


This could be attributed to SAD, short form for [Seasonal Affective Disorder](#), also known as a *seasonal depression* or the *winter depression*.

- People tend to exhibit symptoms of anxiety when there is lesser sunlight ([source](#))

Classification Modelling

- Classification metrics
 - Accuracy (For quick feedback on overall performance of the model making correct predictions - [Hyperparameter Tuning](#))
 - Sensitivity (To be able to minimize missing out identifying users potentially having PTSD, as they could miss out on proper treatment, by the time they've been diagnosed, it might lead to permanent damage if untreated- [Final Model Selection](#))
 - ROC_AUC (Measure of separability between classes - [Final Model Selection](#))



$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{All Positives}} = \frac{TP}{TP + FN} = \frac{TP}{P}$$

Classification Modelling

	iteration	transformer_estimator		GS_best_params	GS_best_score	train_accuracy	test_accuracy
1	Baseline	CountVec LogisticRegression	{'cvec1_max_df': 0.7, 'cvec1_max_features': 1000, 'cvec1_min_df': 0.02, 'cvec1_ngram_range': (1, 1), 'logreg1_max_iter': 1000}		0.810	0.837	0.810
2	Baseline	TfidfVec LogisticRegression	{'logreg2_max_iter': 1000, 'tfvec2_max_df': 0.7, 'tfvec2_max_features': 1000, 'tfvec2_min_df': 0.02, 'tfvec2_ngram_range': (1, 1)}		0.816	0.837	0.816
3	Hyperparameter Tuned	TfidfVec MultinomialNB	{'nb_alpha': 1e-10, 'tfvec_max_df': 0.3, 'tfvec_max_features': 3555, 'tfvec_min_df': 0.0002, 'tfvec_ngram_range': (1, 1)}		0.833	0.857	0.833
4	Hyperparameter Tuned	TfidfVec LogisticRegression	{'logreg4_max_iter': 1000, 'tfvec4_max_df': 0.225, 'tfvec4_max_features': 9500, 'tfvec4_min_df': 0, 'tfvec4_ngram_range': (1, 1)}		0.850	0.901	0.849
5	Hyperparameter Tuned	TfidfVec RandomForest	{'RanForest_max_depth': None, 'RanForest_max_features': 'sqrt', 'RanForest_n_estimators': 300, 'tfvec5_max_df': 0.8, 'tfvec5_max_features': 9000, 'tfvec5_min_df': 0, 'tfvec5_ngram_range': (1, 1)}		0.830	1.000	0.833

Choice of Vectorizer

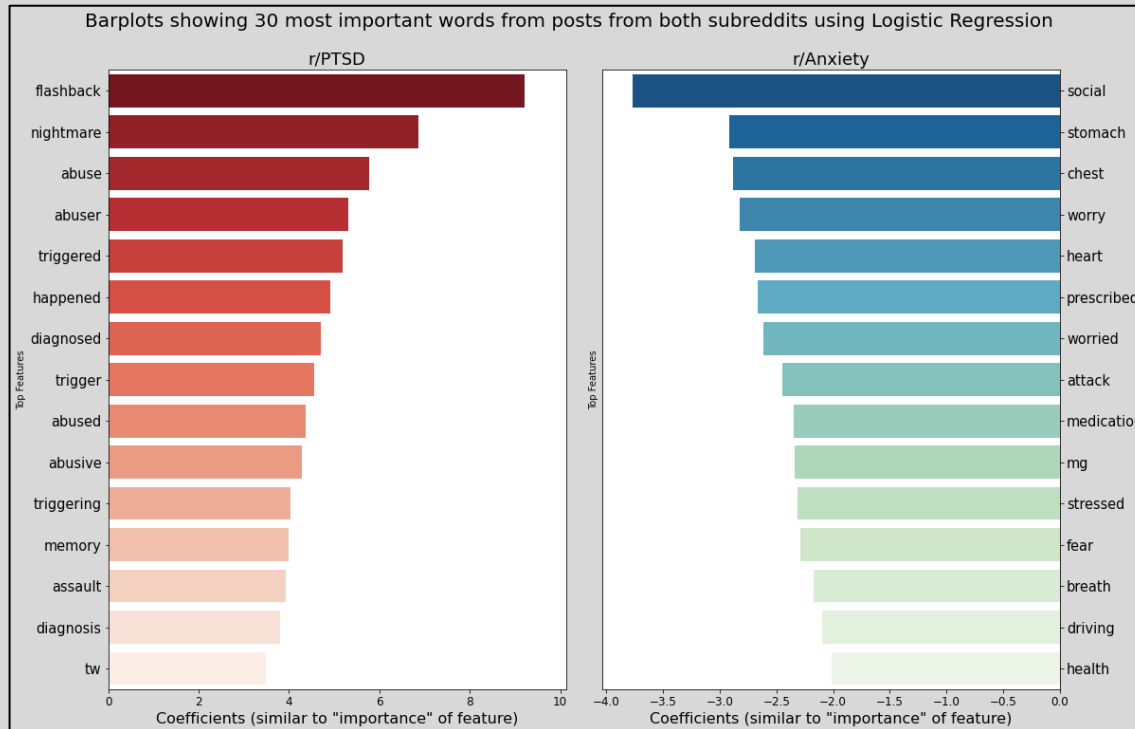
- 1) Chose TfidfVec over CountVec based on Accuracy score and essentially it helps to retrieve words that are rare in some documents that might bring more predictive power.

Key observations

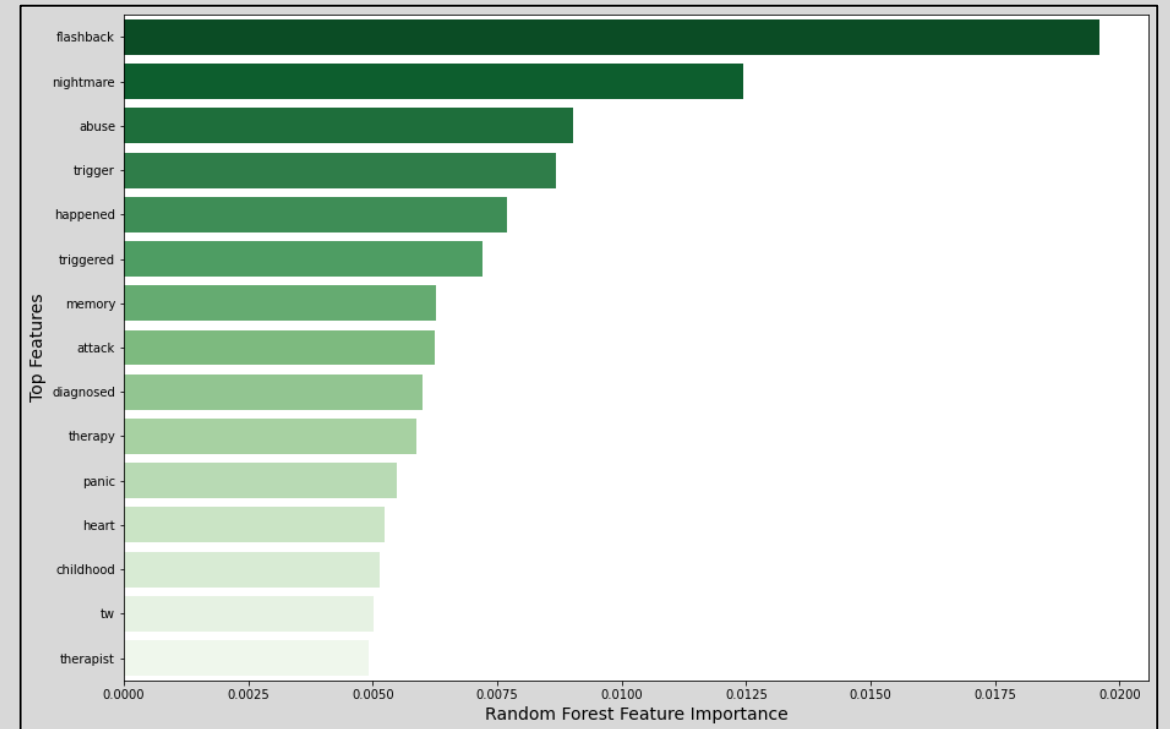
- 1) After using *GridSearchCV*, found that best parameters used is (1,1) indicating there is not much significance in 2-3 word combination that may help with predicting accurately
- 2) RandomForest is highly overfitted, hence unable to generalize to unseen data.
- 3) Although Logistic Regression may have the best accuracy, it tends to overfit as compared to MultinomialNB.

Error Analysis – Feature Importance

Logistic Regression



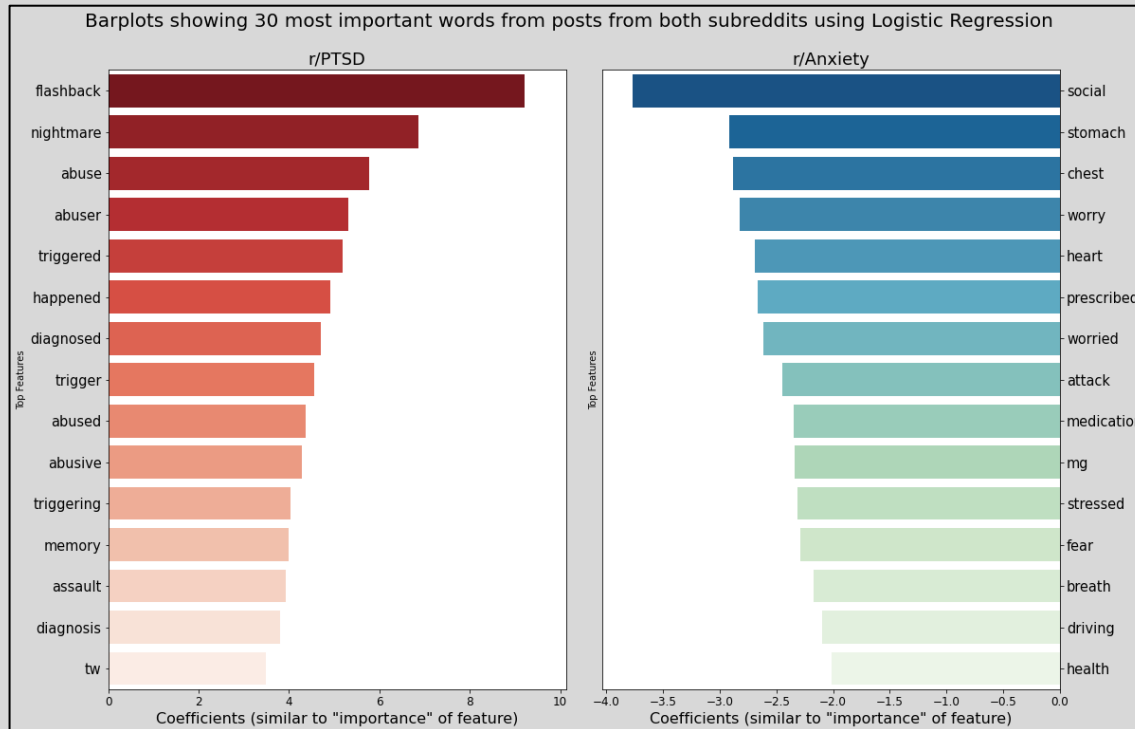
Random Forest



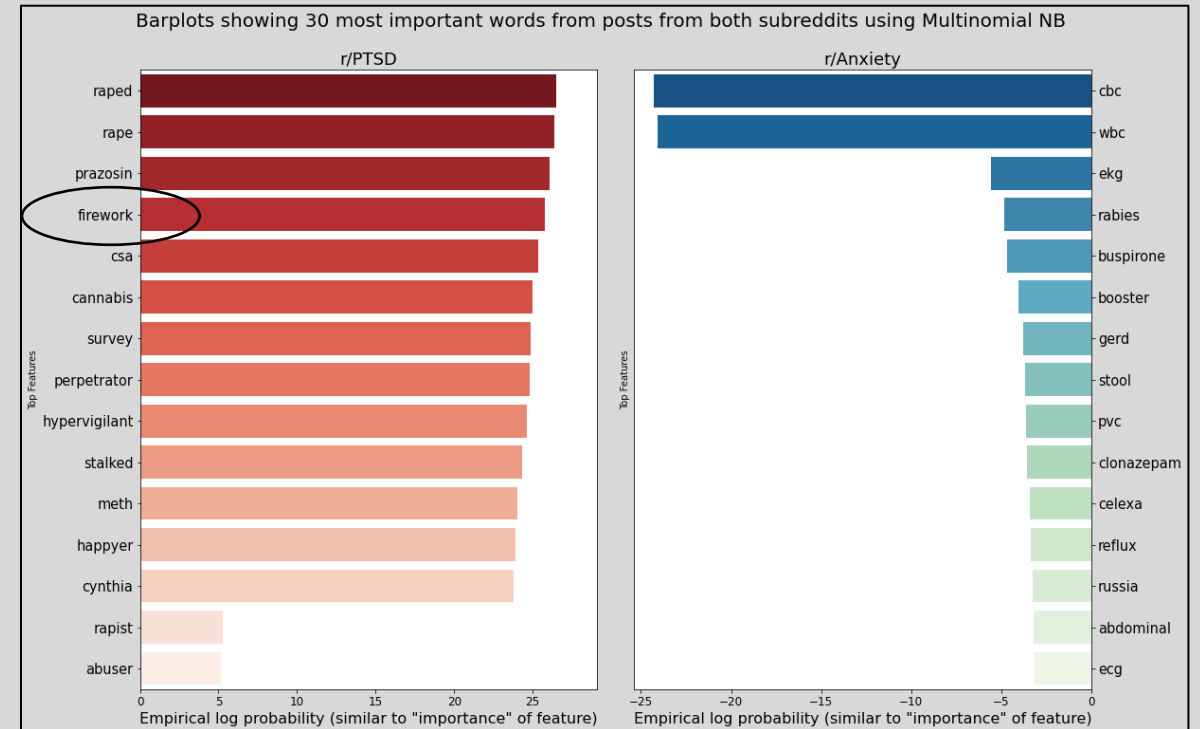
Generally, RandomForest and Logistic Regression has similar top few important features ranked the same way. The features highlights words like *flashbacks*, *nightmares* and *triggers* that can serve as **true symptoms of users suffering from PTSD**

Error Analysis – Feature Importance

Logistic Regression



Multinomial NB



Top few features from MultinomialNB, you can find more words describing the **Causes & Triggers of PTSD** more like *rape*, *"csa"* which stands for *childhood sexual abuse*, *stalked* and *firework*

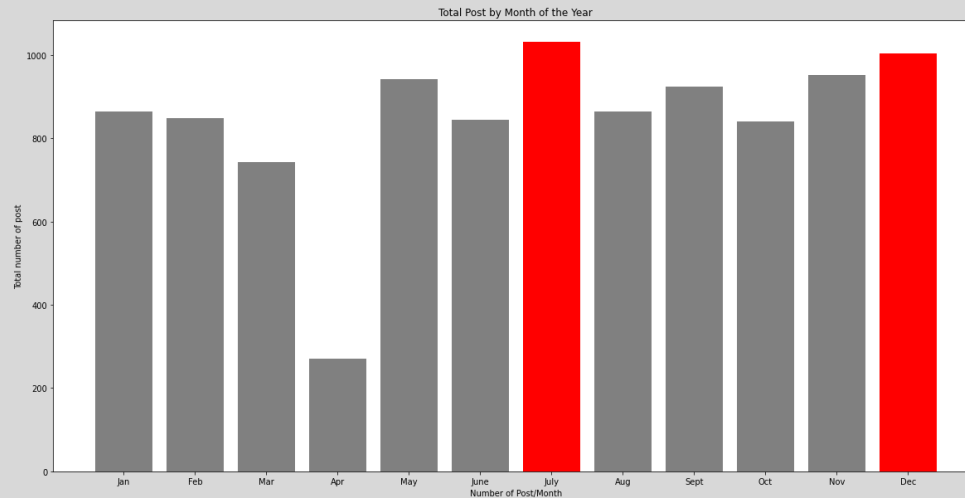
Error Analysis – Feature Importance

Upon investigating the main dataset, users are mainly expressing

- Agony
- Scared
- Triggered thinking it was incoming mortars and rockets.

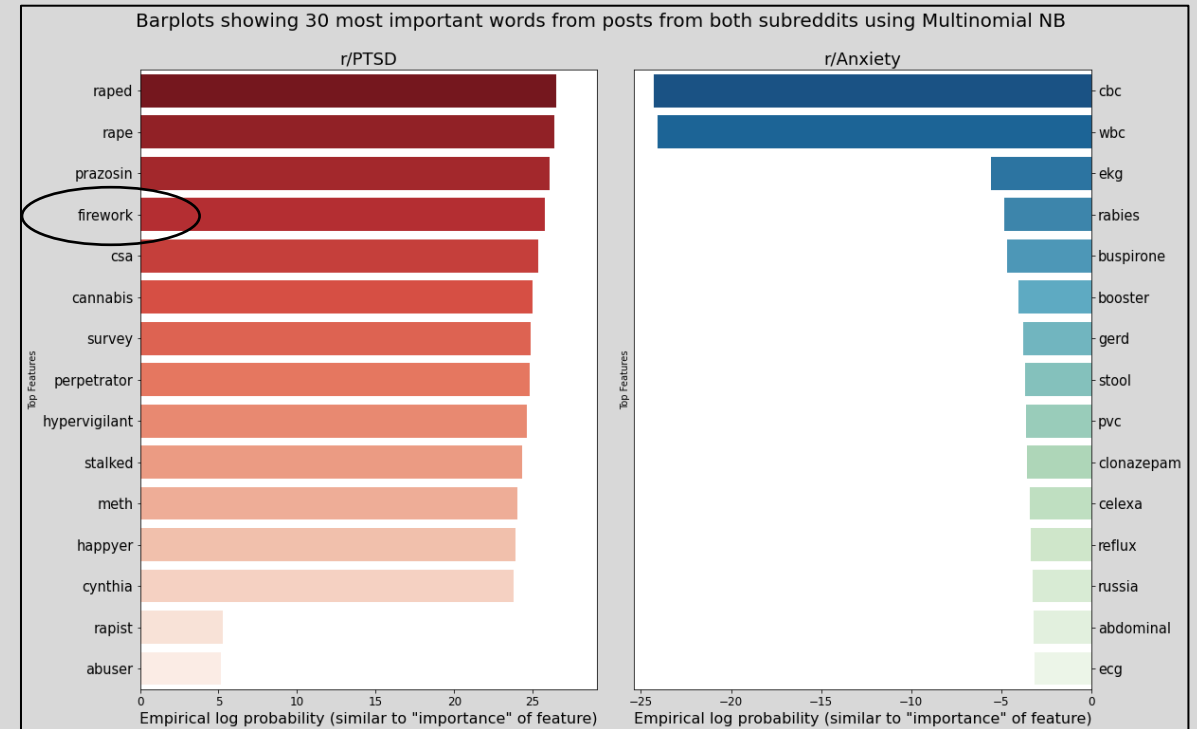
author	
month	
1	7
2	3
5	1
6	7
7	33
8	5
9	4
10	3
11	5
12	2

Number of post in a month



All 70 post about fireworks are from r/PTSD. Based on [source](#), people with past trauma related to sound are impacted worse by loud explosion, reminiscing imminent threat and may activate the brains' threat detectors.

Multinomial NB



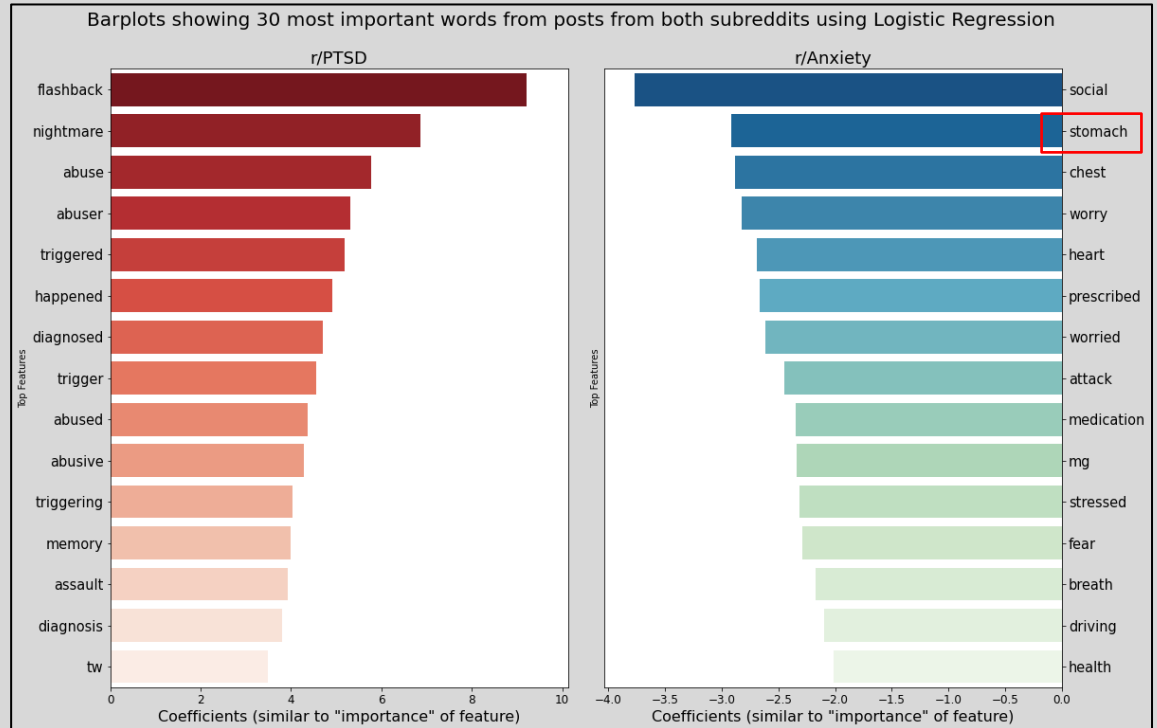
Common mis-classifier

- False **Negative**

Predict : Anxiety, Actual: PTSD

- Missing strong predictive words
- May contain strong words that also describes users in r/Anxiety

Feature	log_prob_diff
stomach	-1.721908
nausea	-1.640161
headache	-1.331461
pain	-0.401746
right	-0.296136
symptom	-0.241098
physical	-0.221189
else	-0.209223
anything	-0.054719
anyone	0.000000
nauseated	0.000000
can	0.000000
not	0.000000
horrible	0.080683
day	0.193628
struggle	0.312020



index	y	y_pred	selftext	lemmi_clean_text
8720	1	0	Does anyone else here struggle with nausea,headaches,stomach pain and so on? Right now im having a horrible day and im so nauseated and off that I cant do anything	anyone else struggle nausea headache stomach pain right horrible day nauseated cannot anything physical symptom

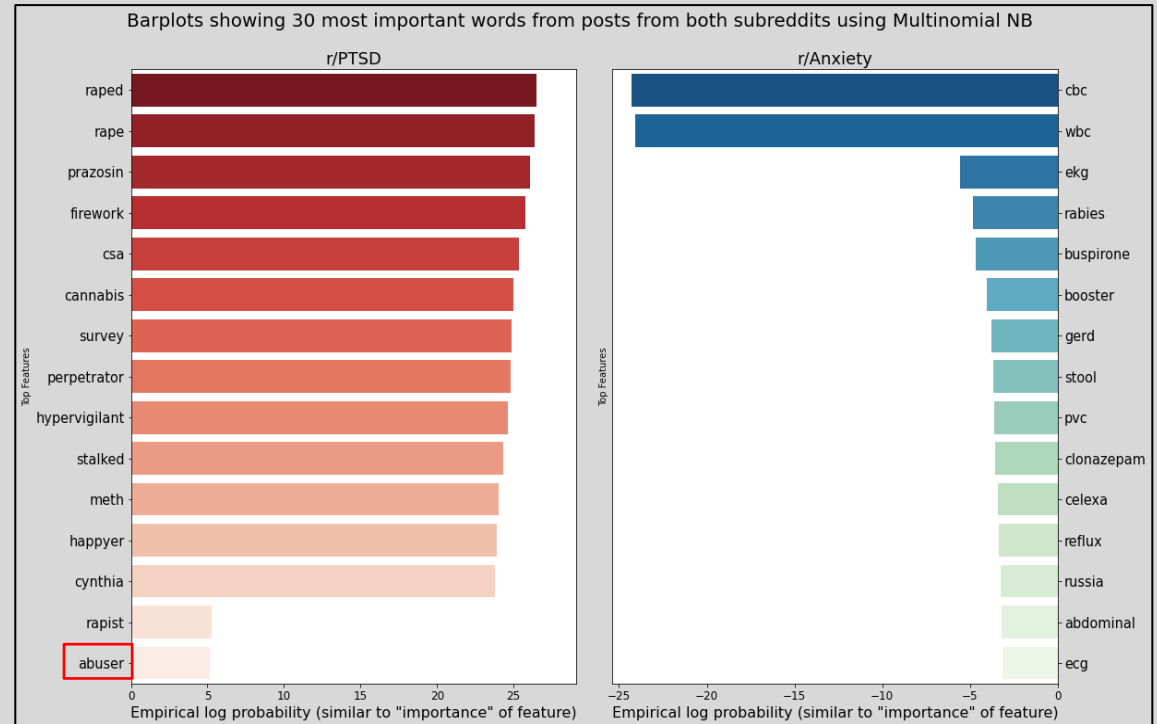
Common mis-classifier

- False Positive

Predicted: PTSD, Actual: Anxiety

- Contains strong features of positive coefficients
- However, **context** was about the father who has PTSD.

Feature	log_prob_diff
abuser	5.157108
veteran	3.511599
abusive	2.363389
abusing	2.339172
military	2.285259
...	...
panic	-0.881445
failure	-0.893583
real	-0.951434
real	-0.951434
attack	-0.975867



y	y_pred	title	selftext	lemmi_clean_text
0	1	My dad says me having anxiety makes me a "p*ssy" and "not a real man"	He also has anxiety and it destroyed his life, but he's projecting onto me after I (M20) developed a panic disorder last year. When he found out how bad it actually was, I thought he would be nice. But he yelled at me and said some horrible things that honestly might put me into a depression for a while. \n\nHe said anxiety is a choice and that since my brain is like plastic, I can choose not to be anxious. And he said I'm gonna be a failure in life if I keep having panic attacks. And he even said he will shove as many triggers in my face whenever possible so I can "get the fuck over it". \n\nHe also told he "I wish someone beat the shit out of you right now. Maybe that would teach you a lesson on manliness." He also said I'm a disgrace to anyone who went through any actual PTSD trauma like marines or veterans.\n\nI told him he's verbally abusing me and he said "I'm telling you this because I love you and want to see you succeed in life! The real world is rougher than I am and I'm trying to prepare you for it! And if this is abusive then I guess every military Sargent out there is an abuser! Haha! See! You can't refute that! You loose!"\n\nCan I please have some support?	also destroyed life projecting onto developed panic last year found bad actually thought would nice yelled said horrible thing honestly might put depression said choice since brain like plastic choose anxious said going failure life keep panic attack even said shove many trigger face whenever possible get fuck also told wish someone beat shit right maybe would teach lesson manliness also said disgrace anyone went actual like marine veteran told verbally abusing said telling love want see succeed life real world rougher trying prepare abusive guess every military sargent abuser haha see cannot refute loose please support dad say make p ssy real man

Common mis-classifier – Users in both subreddits

author	index	y	y_pred	title	selftext	lemmi_clean_text	month	user_contribute_where
yuugyho	4897	0	1	Am I a weak person? Honesty would be appreciated	\nLong story short, I was abused a good chunk of my time in high school and it's still bugging me to this day (I'm 24m). I was regularly abused by the same group of people but the main guy in the group was the one who would physically assault me whenever he got the chance and they were humiliating beat-up's. This means that he was doing wrestling moves on me bc I was so small (he was absolutely massive for a 15 y/o kid), he'd literally hold me in the air with my feet facing the sky and slam me on my stomach while calling me the n word consistently throughout high school. He bullied me on social media as well. He'd always apologize after abusing me and I'd believe him then he'd beat me again and the cycle continued.\n\nHe reached out a while after high school to say he's sorry but I already knew it was a selfish apology to make himself feel better. I accepted it but told him I'm not interested in started any type of friendship and told him to do him while I do me. He was very pleased by my response and wishing me well and all. I've entered insanity and can't grasp the fact I have been abused, no one was listening to me when I tried to get help, and he's all good now like nothing happened after all the suffering. Ive ditched my religion (which was my main source of being forgiving) and I have a heavy heart I'm feeling impulsive. Am I weak for always accepting his apologizes? Was I naive ? How can I be smarter ? I feel extremely insecure about these beatings, so much that I've moved to a different state and deleted all my social media so no one can contact me or see where I'm living.	long story short abused good chunk time high school still bugging day regularly abused group people main guy group one would physically assault whenever got chance humiliating beat-up mean wrestling move bc small absolutely massive kid would literally hold air foot facing sky slam stomach calling n word consistently throughout high school bullied social medium well would always apologize abusing would believe would beat cycle continued reached high school say sorry already knew selfish apology make feel better accepted told interested started type friendship told pleased response wishing well entered insanity cannot grasp fact abused one listening tried get help good like nothing happened suffering ditched religion main source forgiving heavy heart feeling impulsive weak always accepting apologizes naive smarter feel extremely insecure beating much moved different state deleted social medium one contact see living weak person honesty would appreciated	2	both
mpedro12	103	1	0	To Those Who Freeze Up In Social Settings: Have You Found A Way To Overcome It?	\n\nHi there.\n\nMy social anxiety has been very bad lately. Everytime I have a social encounter, my whole body freezes up. I am so fucking nervous that I cant even speak a word. Especially my face muscles are very tense, which means that I cant laugh, smile, or express myself. When I try to talk, I stutter or stumble over my words which is so embarrassing. I avoid eye contact because I feel my vis-a-vis recognizes that I am nervous and stares into my soul. Has anyone found a way to overcome this freezing response, be it medication or any other strategies?	hi social bad lately everytime social encounter whole body freeze fucking nervous cannot even speak word especially face muscle tense mean cannot laugh smile express try talk stutter stumble word embarrassing avoid eye contact feel vis-a-vis recognizes nervous stare soul anyone found way overcome freezing response medication strategy freeze social setting found way overcome	1	both

- False **Positive**

Predicted: PTSD, Actual: Anxiety

- Contains strong features towards positive class like **abused**, **assaulted**.
- User actually posted in Anxiety.

Feature	log_prob_diff
assault	4.544007
abused	3.524986
abused	3.524986
abused	3.524986
abusing	2.341268

- False **Negative**

Predict : Anxiety, Actual: PTSD

- Lacks strong predictive words from positive class
- User actually posted in PTSD.

Feature	log_prob_diff
stutter	-1.998762
social	-1.197056
social	-1.197056
social	-1.197056
nervous	-1.142605

Model Selection

Accuracy improved by 5.8%

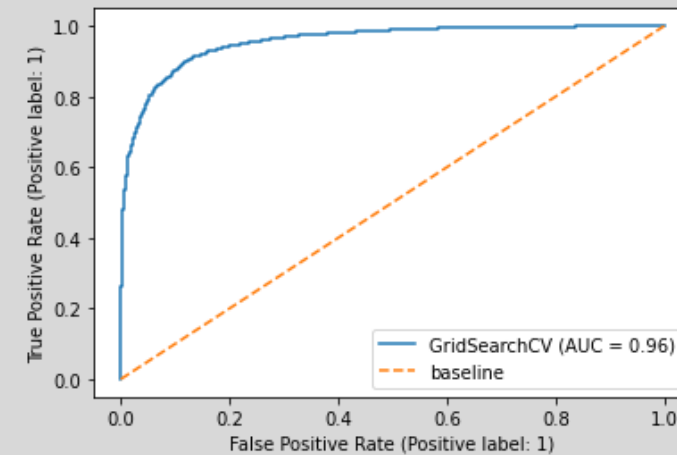
iteration	transformer_estimator	GS_best_params	GS_best_score	train_accuracy	test_accuracy	Sensitivity	Specificity	ROC_AUC
Refined DataSet	TfidfVec LogisticRegression	{'logreg7__max_iter': 1000, 'tfvec7__max_df': 0.3, 'tfvec7__max_features': 8375, 'tfvec7__min_df': 0, 'tfvec7__ngram_range': (1, 2)}	0.848	0.904	0.843	0.860925	0.907759	0.956325
Hyperparameter Tuned	TfidfVec LogisticRegression	{'logreg4__max_iter': 1000, 'tfvec4__max_df': 0.22, 'tfvec4__max_features': 10000, 'tfvec4__min_df': 0, 'tfvec4__ngram_range': (1, 1)}	0.845	0.896	0.844	0.815093	0.872526	0.924666
Refined Dataset	TfidfVec MultinomialNB	{'nb_alpha': 1e-10, 'tfvec__max_df': 0.3, 'tfvec__max_features': 3700, 'tfvec__min_df': 0.0, 'tfvec__ngram_range': (1, 1)}	0.834	0.858	0.834	0.864085	0.832937	0.924117
Hyperparameter Tuned	TfidfVec MultinomialNB	{'nb_alpha': 1e-10, 'tfvec__max_df': 0.3, 'tfvec__max_features': 3600, 'tfvec__min_df': 0.0002, 'tfvec__ngram_range': (1, 1)}	0.828	0.854	0.827	0.846306	0.807997	0.908489
Hyperparameter Tuned	TfidfVec RandomForest	{'RanForest__max_depth': None, 'RanForest__max_features': 'sqrt', 'RanForest__n_estimators': 300, 'tfvec5__max_df': 0.8, 'tfvec5__max_features': 7000, 'tfvec5__min_df': 0, 'tfvec5__ngram_range': (1, 1)}	0.825	1.000	0.830	0.802450	0.857878	0.905106
Baseline	TfidfVec LogisticRegression	{'logreg2__max_iter': 1000, 'tfvec2__max_df': 0.7, 'tfvec2__max_features': 1000, 'tfvec2__min_df': 0.02, 'tfvec2__ngram_range': (1, 1)}	0.809	0.830	0.810	0.772027	0.847981	0.892497
Baseline	CountVec LogisticRegression	{'cvec1__max_df': 0.7, 'cvec1__max_features': 1000, 'cvec1__min_df': 0.02, 'cvec1__ngram_range': (1, 1), 'logreg1__max_iter': 1000}	0.801	0.826	0.803	0.753852	0.853127	0.883356

Logistic regression

- For its best accuracy and sensitivity score.

ROC Score

- The better the score, the better it is at distinguishing patients with PTSD and Anxiety.



Conclusion & Recommendations



Problem statement is to provide social service a helping hand in accurately identifying people having PTSD from Anxiety

- ✓ Phase 1: Used subreddit to obtain sufficient documents to start off a model.
- ✓ Phase 2: Hyperparameter tuned and refined dataset to improve model
 - Still have room for improvement to increase metrics like ROC_AUC , Accuracy and Sensitivity
- ❑ Phase 3: Implement to social services incorporate in
 - Entry forums for new patients who wishes to go for counselling (to collect more data)
 - Used for complementary support platform (rather than a main system) to help prompt social workers early.
 - Main idea is to continue collecting real time context because with professional feedback on jargons used and proper classification, it should gain more accuracy and better predicting power while already implementing this to help classify people having PTSD with considerable accuracy.
 - Keeping in mind having high peaks of the number of post in July and Dec, you may consider running a campaign to beef up support for people and use these high peaks as a matrix in the following years to track the effectiveness of the campaigns.