Multi-Class Classification of Mental Health Disorders

Presented by Team 10 - Boys
Aayush Acharya
Deeptansh Sharma

Problem Statement and Requirements

This project deals with the classification of mental health disorders from the Reddit Mental Health dataset.

01.

- Implement the following models from scratch and improve their accuracy as much as possible
  - Random Forest Classifier
  - CNN
  - RNN
  - distilBERT (only fine-tuning)

02. Provide an intuitive and explanatory analysis of the results.

03. Provide an analytical pipeline to differentiate and explain the weaknesses of each model

What is the dataset?

Dataset used is Reddit Mental Health Dataset - https://zenodo.org/records/3941387

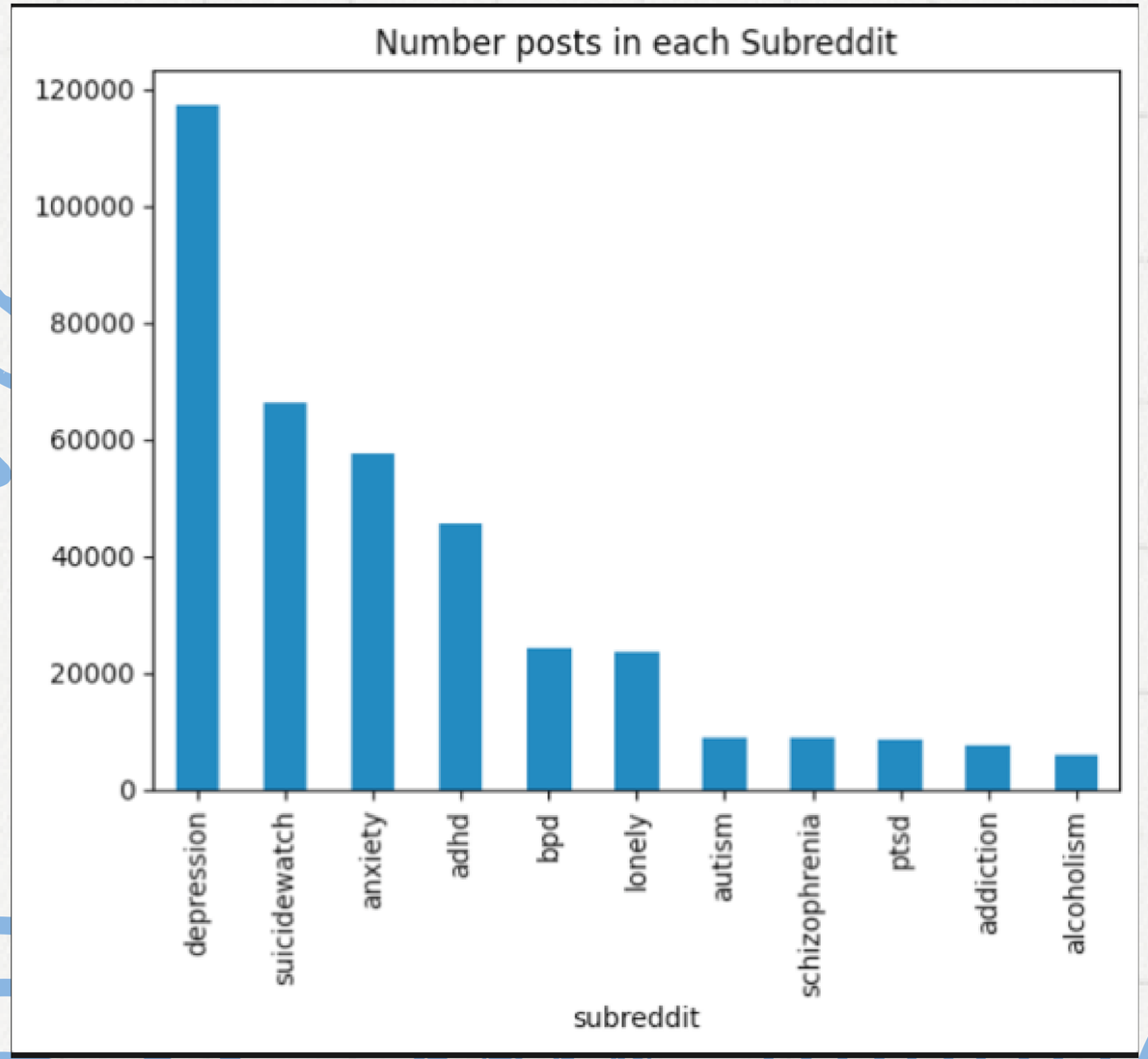
This dataset contains posts from 28 subreddits (15 mental health support groups) from 2018-2020.
Contains posts and text features for the following timeframes from 26 mental health and non-mental health subreddits:

- 15 specific mental health support groups
- 11 non-mental health subreddits

We dropped the non-mental health subreddits as they were not relevant to the mental health disorder classification task.

Dataset Analysis

Distribution of posts in each subreddit



An analysis reveals significant class imbalance towards ‘depression’ and ‘suicidewatch’, with depression having a post count of 117000.
The minority class - alcoholism only has a post count of 6000

Dataset Analysis

Calculated correlation with target

	Column	Correlation
346	subreddit_encoded	1.000000
17	sent_neg	0.257640
301	tfidf_suicid	0.257270
206	tfidf_kill	0.243008
145	tfidf_die	0.204068
328	tfidf_want	0.183635
109	tfidf_anymor	0.173392
46	liwc_death	0.172362
214	tfidf_life	0.171300
217	tfidf_live	0.168835
79	liwc_sadness	0.147013
157	tfidf_end	0.132902
26	suicidality_total	0.131850
177	tfidf_fuck	0.128028
81	liwc_sexual	0.127229
176	tfidf_friend	0.127179
131	tfidf_care	0.121360
166	tfidf_famili	0.114730
22	isolation_total	0.110676
52	liwc_friends	0.110039
84	liwc_swear_words	0.107678
34	liwc_anger	0.106134
244	tfidf_noth	0.102788
200	tfidf_hurt	0.100715
...		
107	tfidf_anxieti	-0.232619
8	wiener_sachtextformel	-0.238629
18	sent_neu	-0.287691
94	tfidf_adhd	-0.347325

We computed the correlation of features with the target variable, which turned out to be quite low for most of the features.
This brings in the need to generate features of our own

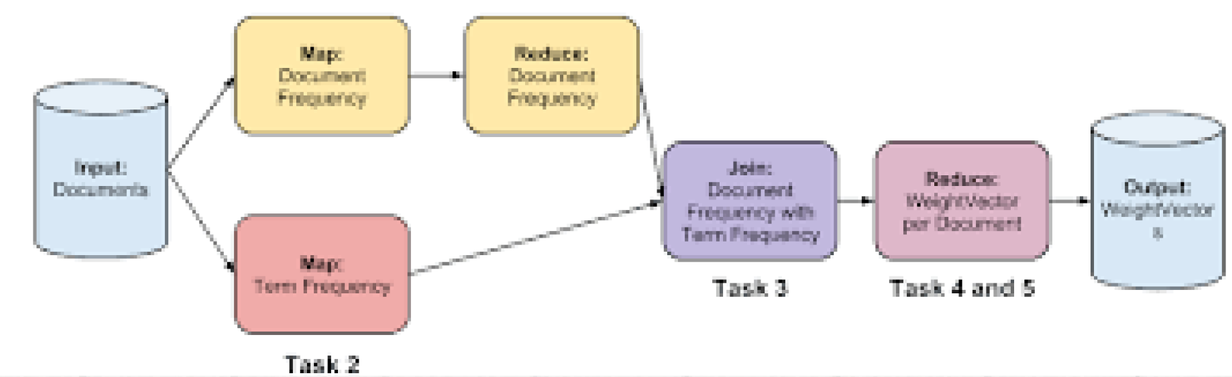
We then computed the ANOVA F_statistic score for the features for a clearer picture

Feature Generation : TF-IDF

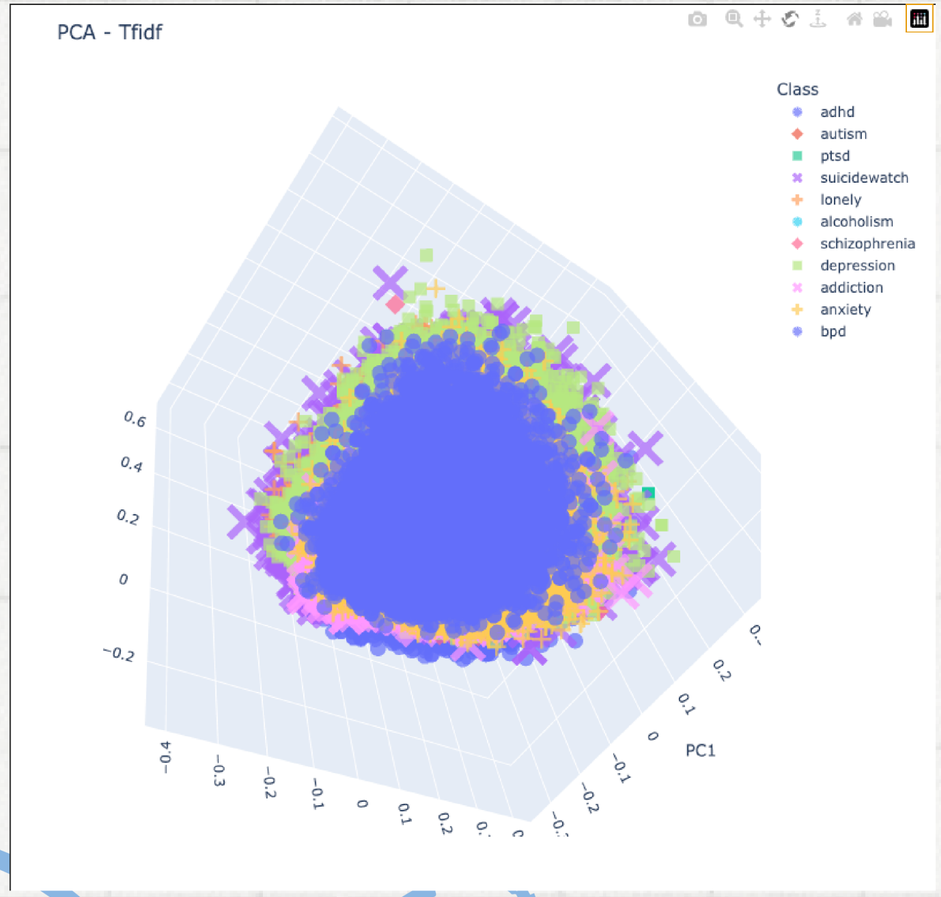
What is it?

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure used to evaluate the importance of a term within a document in a corpus.

In essence, TF-IDF helps in identifying important words or phrases within a document by considering both their local importance (within the document) and their global rarity (across the corpus).



Feature Generation : TF-IDF



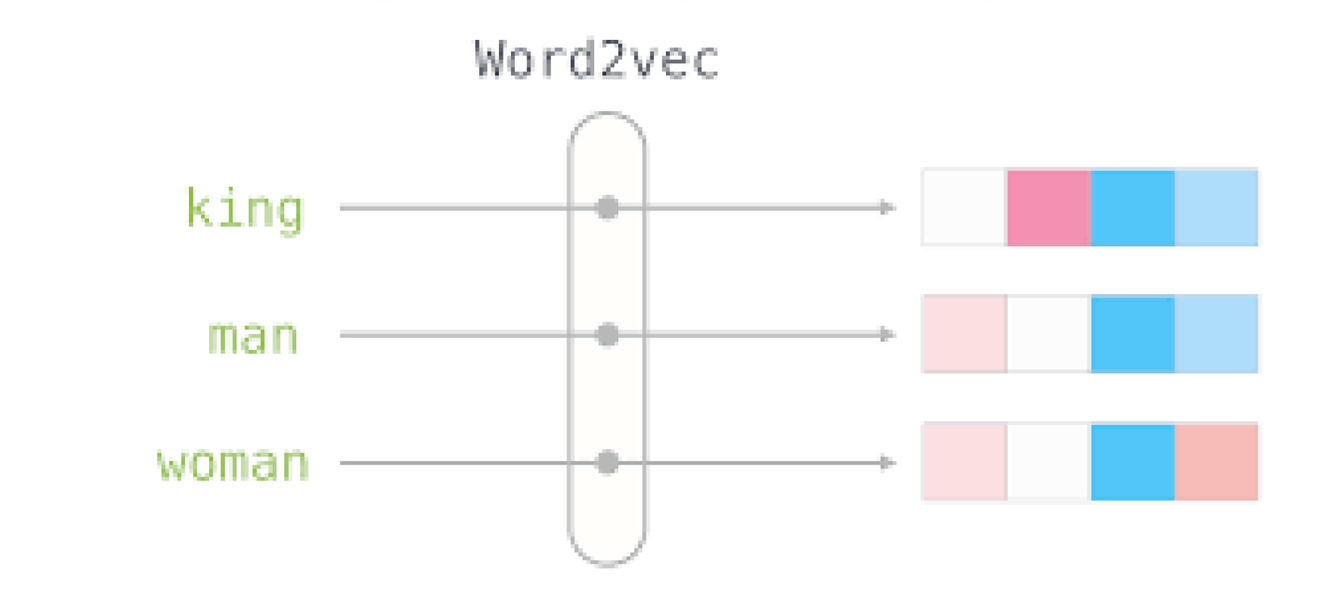
We generated TF-IDF scores for the ‘post’ data and plotted the variance distribution against the class labels

Feature Generation : Word2Vec

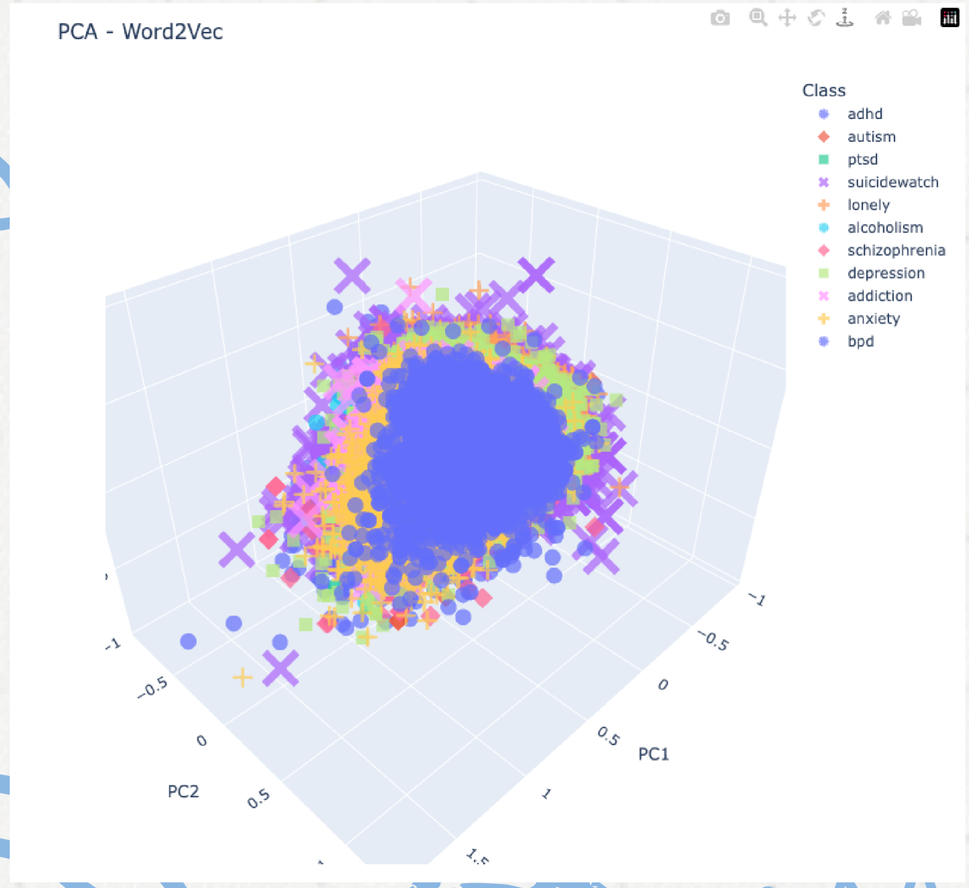
What is it?

Word2Vec is a type of word embedding technique that converts words or phrases from a vocabulary into high-dimensional vectors of real numbers.

It provides dense and meaningful word representations that capture semantic similarities between words, making it highly desirable as a feature for the task



Feature Generation : Word2Vec

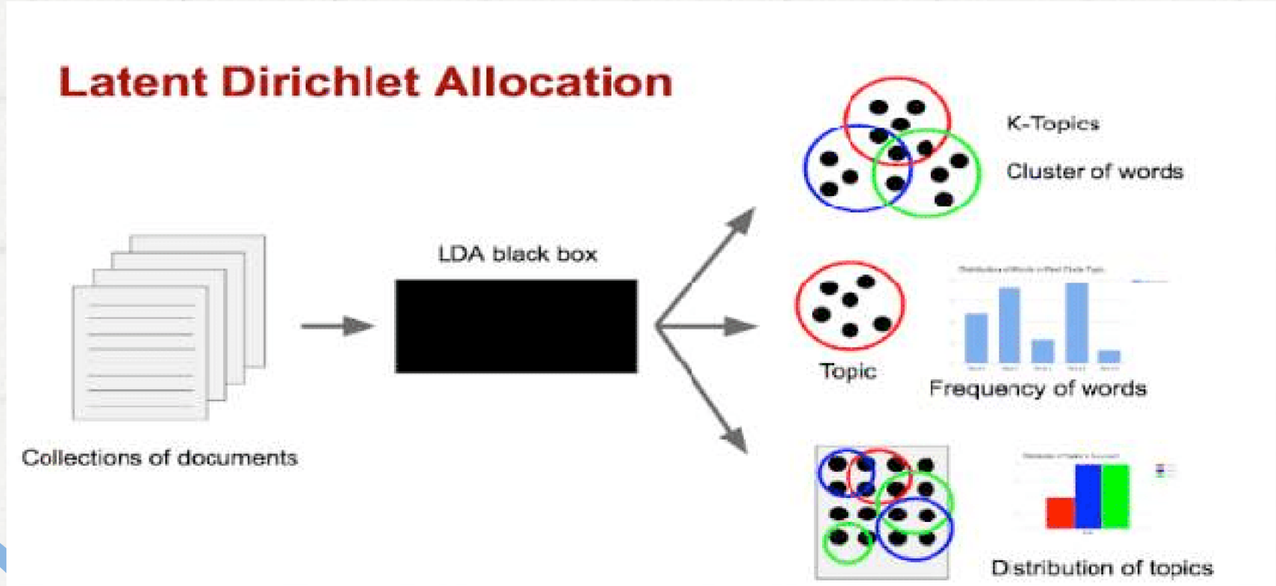


We generated word2vec embeddings of our data and plotted the variance distrbution

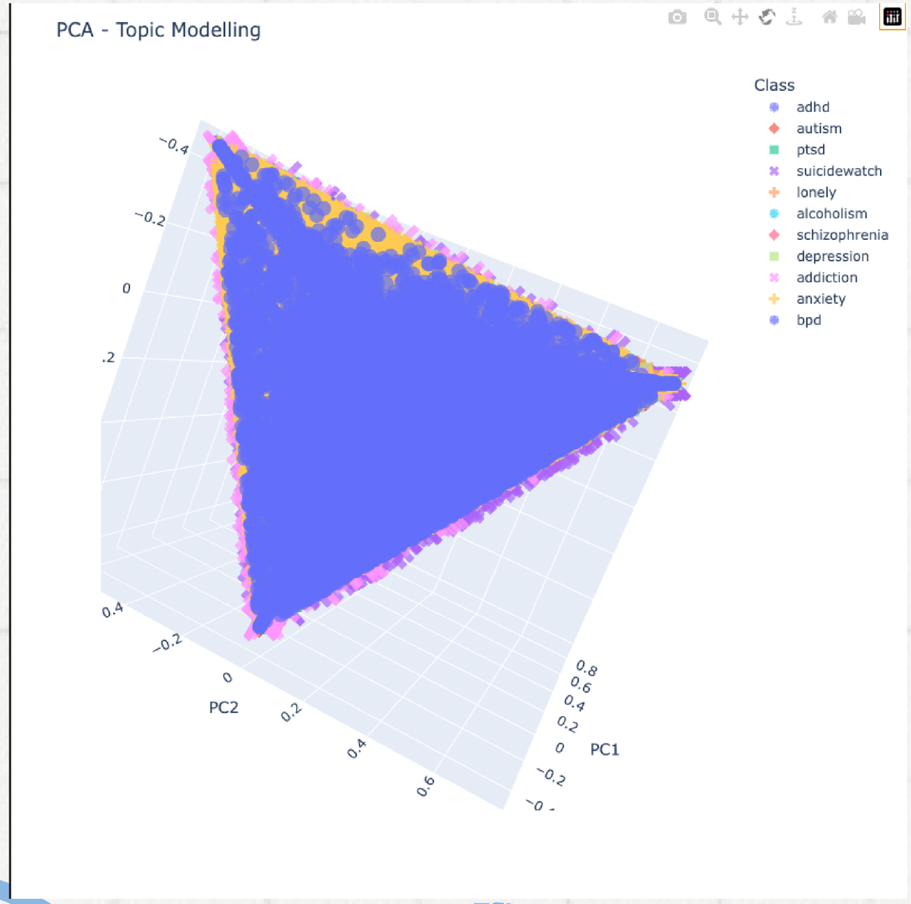
Feature Generation : Topic Modelling

What is it?

Topic Modelling using Latent Dirichlet Association (LDA) is a statistical modelling technique designed to uncover hidden thematic structure within a collection of documents.



Feature Generation : Topic Modelling



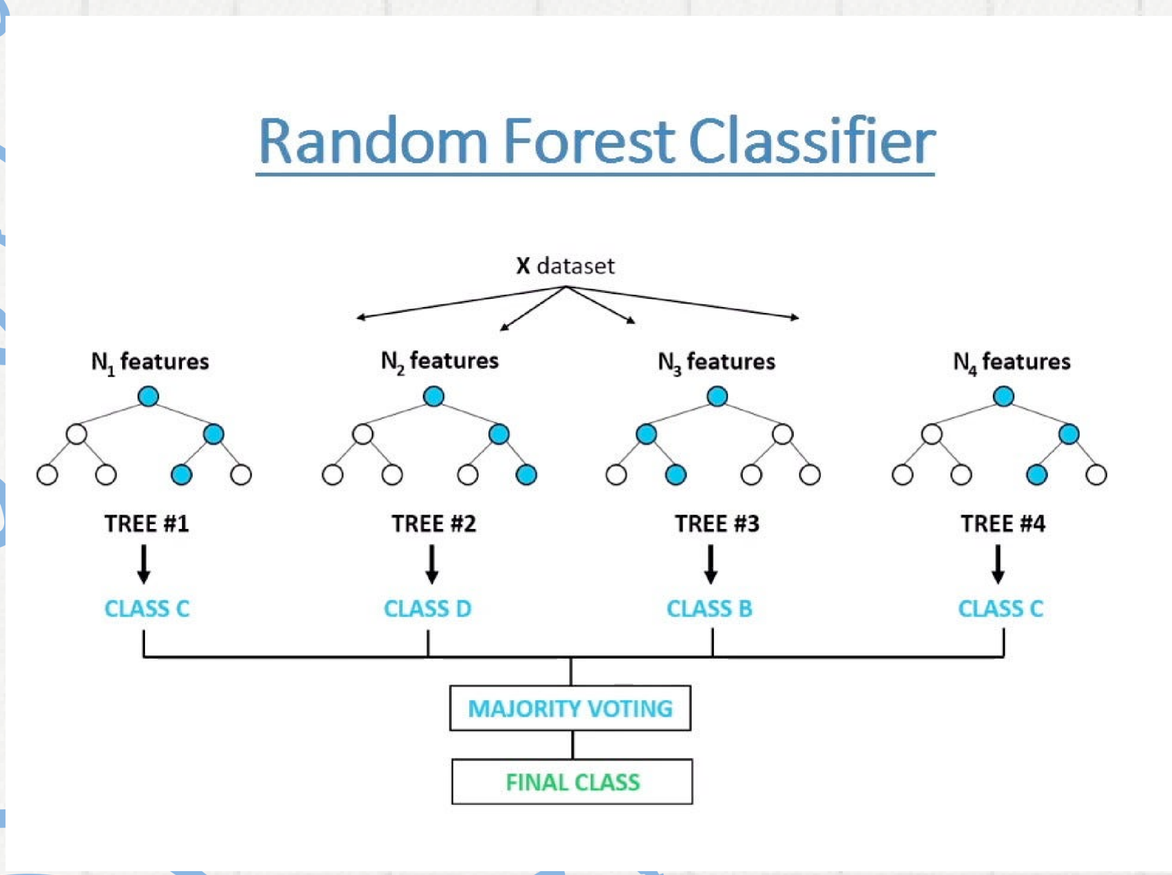
We decided to use LDA as the posts themselves are divided into topics and plotted the variance distribution of the features.

However, it did not give us the results we were expecting , and decided to not use it for training our models.

Models

Random Forest Classifier

What is it?



Random Forest is part of the ensemble learning methods, which combines multiple individual models to produce a more robust and accurate final model. In this case, it combines multiple decision trees to make predictions.

Random Forest Classifier

Setting the baseline

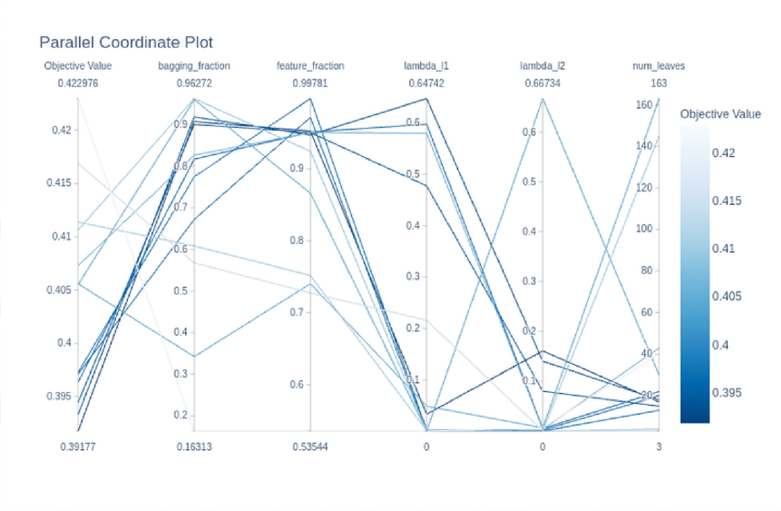
We first ran the inbuilt sklearn model to get baseline numbers.

- Ran it on the original data to get a base result.
- Ran it on the individual features we generated
- Finally ran it on the combined data of original data and generated features

Sklearn RFC				
Type of model	Accuracy	Precision (macro avg)	Recall (macro avg)	F-1 score (macro avg)
word2vec	0.606	0.76	0.51	0.58
Data	0.664	0.79	0.56	0.63
Tfidf	0.684	0.78	0.65	0.70
Generated Features + Data (reduced)	0.682	0.82	0.62	0.69

Random Forest Classifier

Hyperparameter Tuning



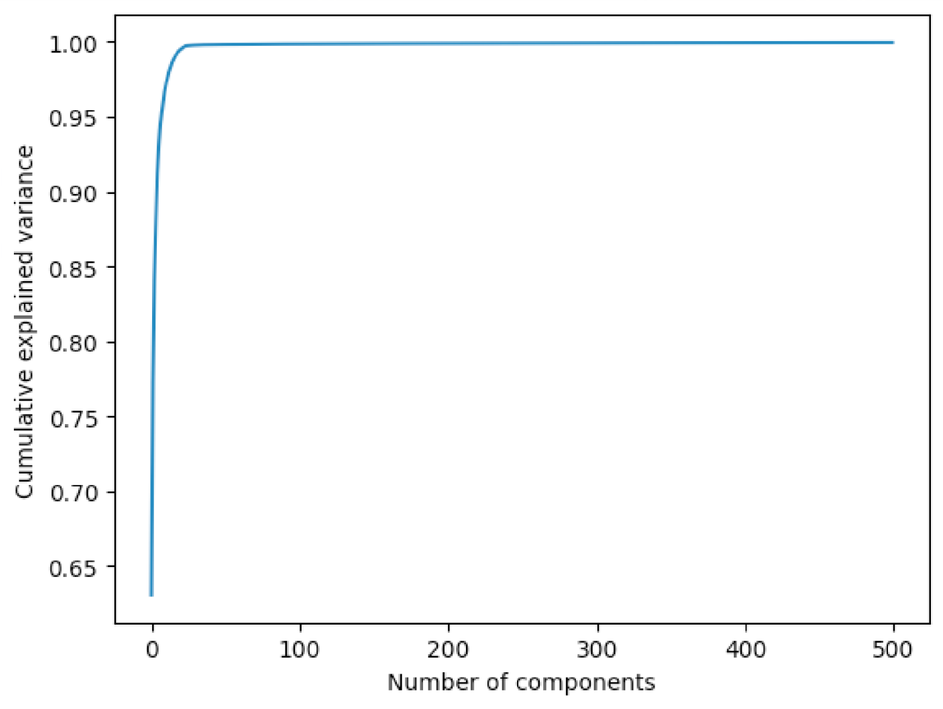
We tuned on the hyperparameters using the OPTUNA Library and got our best hyperparameters.

```
Trial 14 finished with value: 0.6277322143810795 and parameters: {'n_estimators': 349, 'max_depth': 24}. Best is trial 14 with value: 0.6277322143810795.
Trial 15 finished with value: 0.6164562273507533 and parameters: {'n_estimators': 328, 'max_depth': 21}. Best is trial 14 with value: 0.6277322143810795.
Trial 16 finished with value: 0.5885714285714285 and parameters: {'n_estimators': 202, 'max_depth': 18}. Best is trial 14 with value: 0.6277322143810795.
Trial 17 finished with value: 0.6224909403013542 and parameters: {'n_estimators': 357, 'max_depth': 43}. Best is trial 14 with value: 0.6277322143810795.
Trial 18 finished with value: 0.6282586305550257 and parameters: {'n_estimators': 428, 'max_depth': 25}. Best is trial 18 with value: 0.6282586305550257.
Trial 19 finished with value: 0.5581613580011444 and parameters: {'n_estimators': 319, 'max_depth': 15}. Best is trial 18 with value: 0.6282586305550257.
Trial 20 finished with value: 0.6232920083921419 and parameters: {'n_estimators': 417, 'max_depth': 42}. Best is trial 18 with value: 0.6282586305550257.
Trial 21 finished with value: 0.6279610909784474 and parameters: {'n_estimators': 449, 'max_depth': 27}. Best is trial 18 with value: 0.6282586305550257.
Trial 22 finished with value: 0.6277398436009918 and parameters: {'n_estimators': 377, 'max_depth': 24}. Best is trial 18 with value: 0.6282586305550257.
Trial 23 finished with value: 0.467427045584589 and parameters: {'n_estimators': 439, 'max_depth': 5}. Best is trial 18 with value: 0.6282586305550257.

Trial 26 finished with value: 0.5591836734693878 and parameters: {'n_estimators': 397, 'max_depth': 15}. Best is trial 24 with value: 0.6287926759488842.
Trial 27 finished with value: 0.6268319664314324 and parameters: {'n_estimators': 286, 'max_depth': 29}. Best is trial 24 with value: 0.6287926759488842.
Trial 28 finished with value: 0.6230097272553882 and parameters: {'n_estimators': 500, 'max_depth': 38}. Best is trial 24 with value: 0.6287926759488842.
Trial 29 finished with value: 0.6220484455464429 and parameters: {'n_estimators': 214, 'max_depth': 49}. Best is trial 24 with value: 0.6287926759488842.
```

Random Forest Classifier

Building our model



Due to computational constraints, we decided to run PCA on our data and plot the cumulative explained variance plot.
From this, we decided to keep the number of components as 70

Random Forest Classifier

Building our model

We built a Random Forest Classifier from scratch and ran it on our data

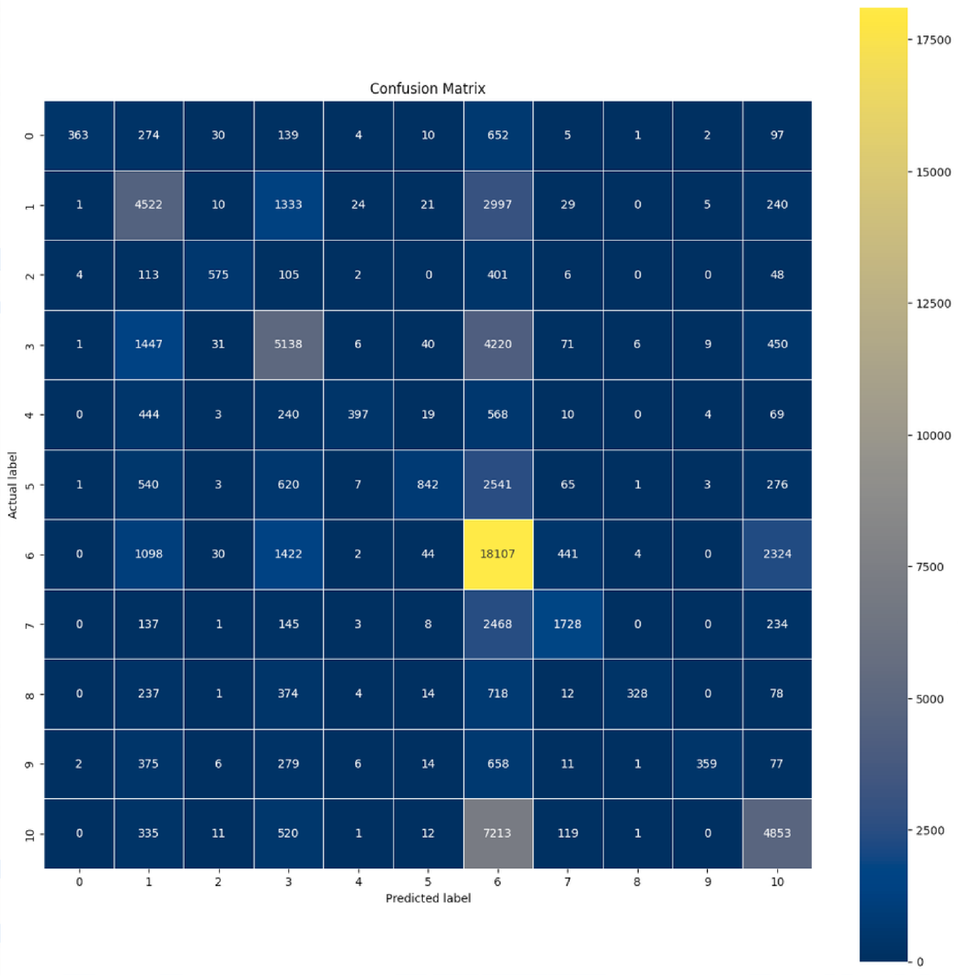
- Ran it on the original data
- Ran it on our data + generated features

RFC from scratch				
Type of model	Accuracy	Precision (macro avg)	Recall (macro avg)	F-1 score (macro avg)
Data + PCA(70)	0.496	0.73	0.36	0.43
Data (reduced) + Generated Features + PCA(70)	0.621	0.71	0.54	0.59

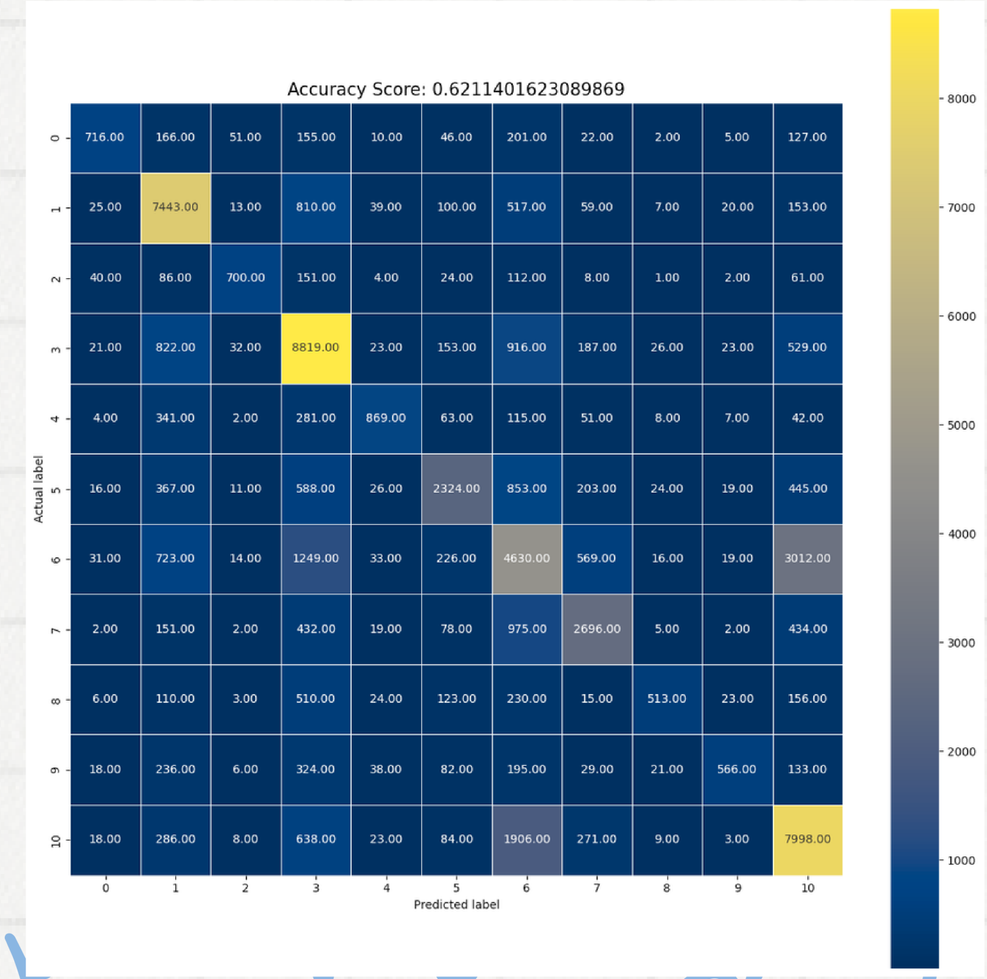
Random Forest Classifier

Results

Original Data



With generated features



Random Forest Classifier

Analysis

Advice life Hello I diagnosed 18 I currently 20 I started studying computer science since starting wonder I never special interest I usually feel happy I find I feel least happy exam stressful time like I sure occupation computer scientist match I always thought I would happy detective Is possible people like u get bureaucracy stress university

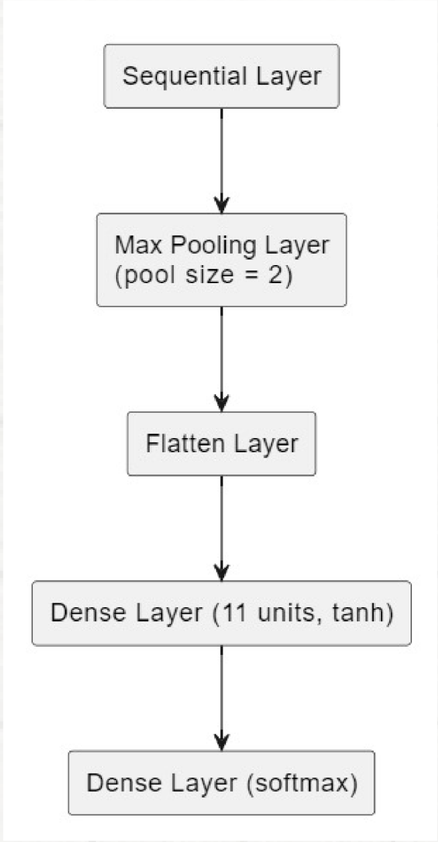
Actual Class - ADHD

- Handling Sequential Information: RFC doesn't inherently consider the sequential nature of data, like text embeddings represented by Word2Vec. Word order and context are crucial in text analysis.

- Dimensionality and Sparse Data: Word2Vec embeddings and TF-IDF scores often result in high-dimensional, sparse data. RFC might struggle with high dimensionality and sparse feature spaces, potentially leading to increased computational complexity and a risk of overfitting.

Convolutional Neural Network

What is it?



CNN's are neural networks that are primarily used for analyzing visual imagery.They consist of multiple layers, primarily convolutional layers. These layers perform convolutions (mathematical operations) on the input image using learnable filters or kernels.

Convolutional Neural Network

Building and running our model

We built our model and trained it in the following ways:

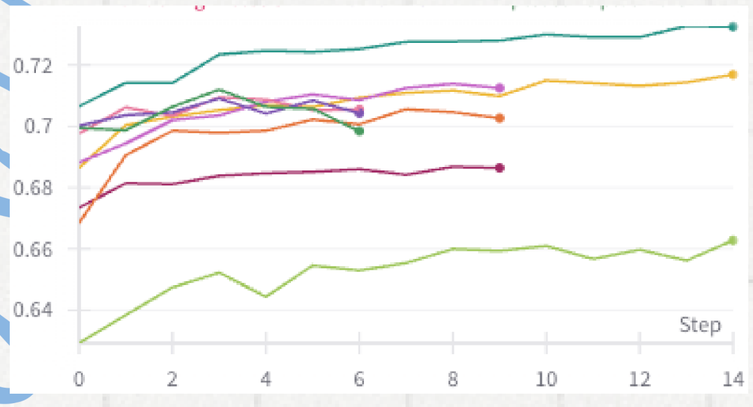
- Ran it on the original data to get a base result.
- Ran it on the combined data of original data and generated features

CNN				
Type of data	Accuracy	Precision (macro avg)	Recall (macro avg)	F-1 score (macro avg)
Data	0.668	0.68	0.67	0.67
Generated Features + Reduced Data	0.712	0.76	0.68	0.72

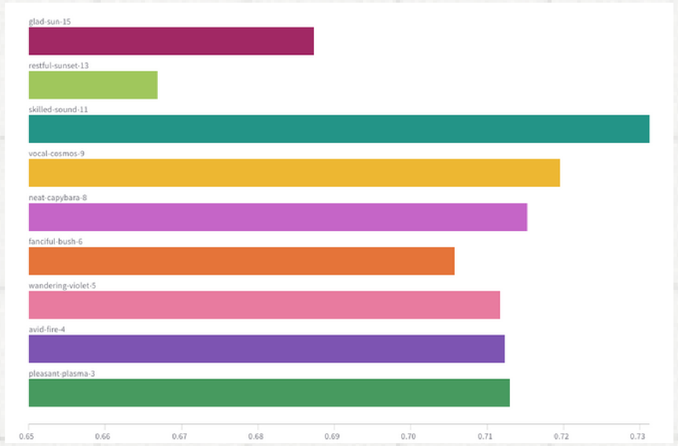
Convolutional Neural Network

Hyperparameter Tuning

Validation Accuracy



Test Accuracy



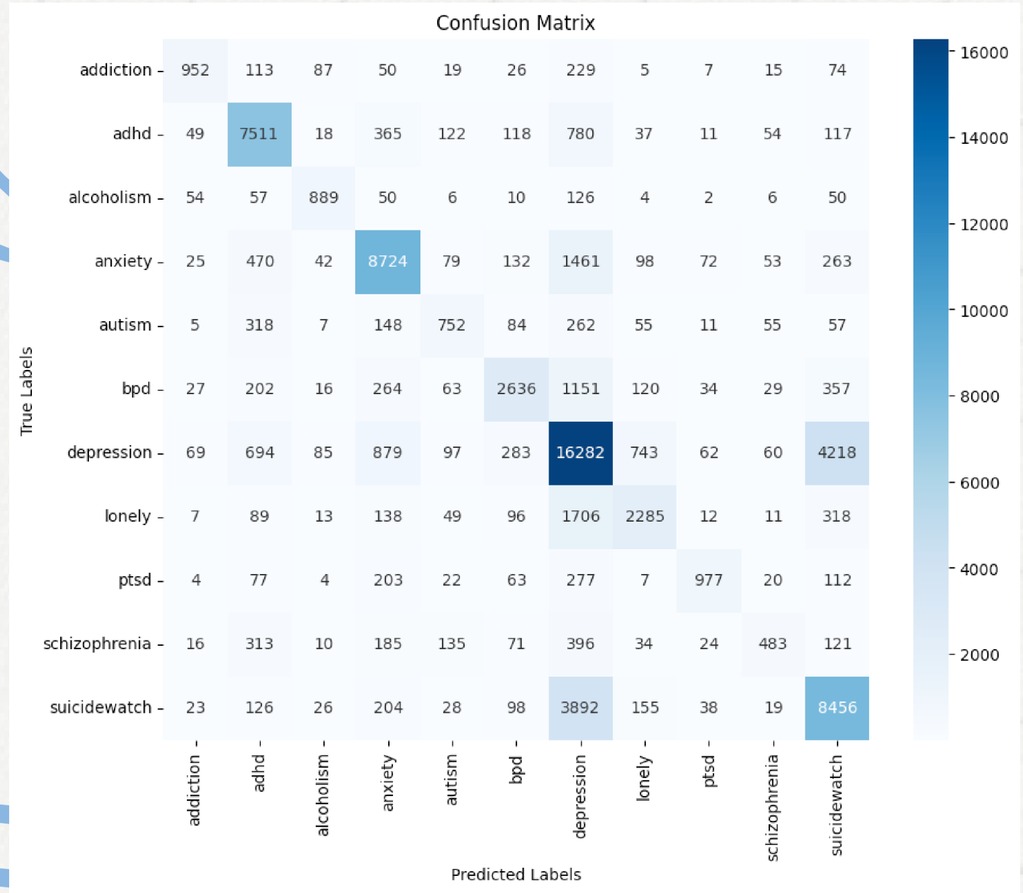
We tuned our model on the following hyperparameters:

- Filters : [**16**, 32]
- Kernel size: [2, **3**, 6]
- Pool size: [**2**,3]
- Dense Units: [100, 150, **200**, 300]
- Activation function: [**tanh**, relu]

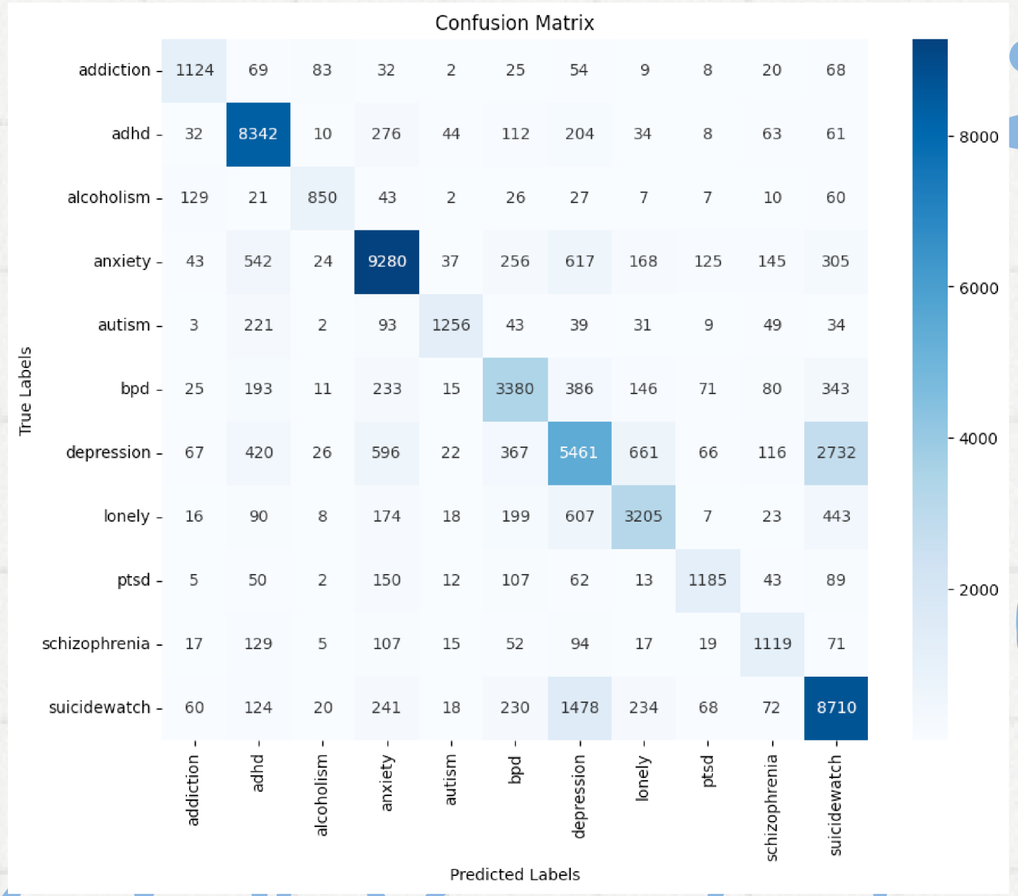
Convolutional Neural Network

Results

Original Data



With generated features



Convolutional Neural Network

Analysis

Actual Label - PTSD

Hallucinations occurring frequently NSFW Marked NSFW due potentially triggering content Just little backstory I suffered PTSD since I 10 year old I nearly 20 diagnosed PTSD I 18 For year I would nightmare I would relive experience seems pretty common When I got 14 I would go 2 3 day without sleeping avoid nightmare I know cope When I would go 2 3 day without sleeping I would often hear laughter incoherent whisper real sound someone hitting metal object next head I alone It pretty easy write result sleep deprivation go day even I found horrifying In recent year nightmare become lot sporadic occur often I cope nightmare next day meditation happen Because I gone several day without sleeping year I free hallucination quite time The dilemma I feel lot serious I deprived sleep About month two ago I experienced one flashback nightmare I almost decade Early next day I began meditating Everything going planned 20 minute At point I hearing man speaking aggressively French much like Jacques Brel voice sounded deeper I fluent French whatever I hearing mystery The next day similar thing happened nightmare except time sounded like airplane head It loud enough scare meditation make heart race The nightmare le frequent since I however experiencing straightforward hallucination Every hour I see humanoid figure corner eye Like someone behind watching More recently I also seeing figure running around right front The night I girlfriend room watching show I noticed corner room shadowy figure corner eye I could focus enough make feature I best make sure I really saw something When I finally glanced gone She knew something I seemed distracted maybe sensed terror asked wrong I said I thought I saw something I afraid come forward anyone everyday life I afraid people may think something wrong I need someone verify normal PTSD another mental illness I also appreciate anecdote advice Thank advance This scary I felt like I well long

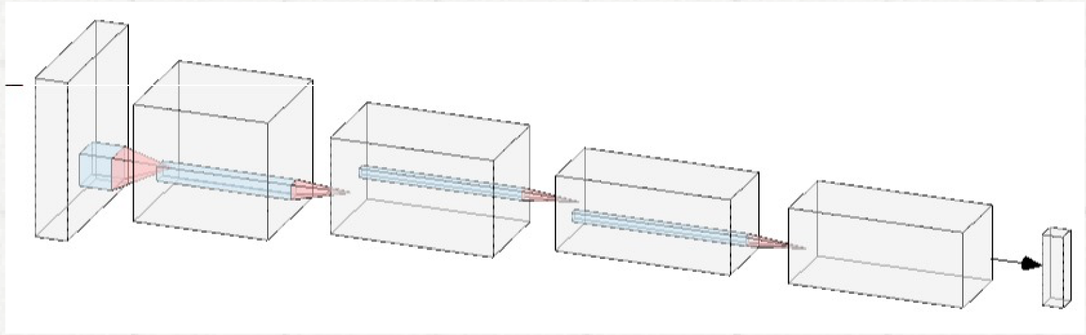
TF-IDF Scores: CNNs might not effectively utilize TF-IDF scores as features. TF-IDF scores highlight the importance of words in a document based on their frequency and rarity across the corpus, but CNNs might not inherently grasp the significance of these weighted scores in the convolutional operations.

Fixed Window Size: CNNs typically use a fixed-size window to perform convolutions, which might not effectively capture the varying lengths of sentences in the text data. This can result in a loss of information or context from longer sentences.

Data Representation: Word2Vec embeddings might not fully capture the intricacies of the text data. While they represent words in a dense vector space, some nuanced semantic information might be lost or might not be effectively utilized by the CNN architecture.

Recurrent Neural Network

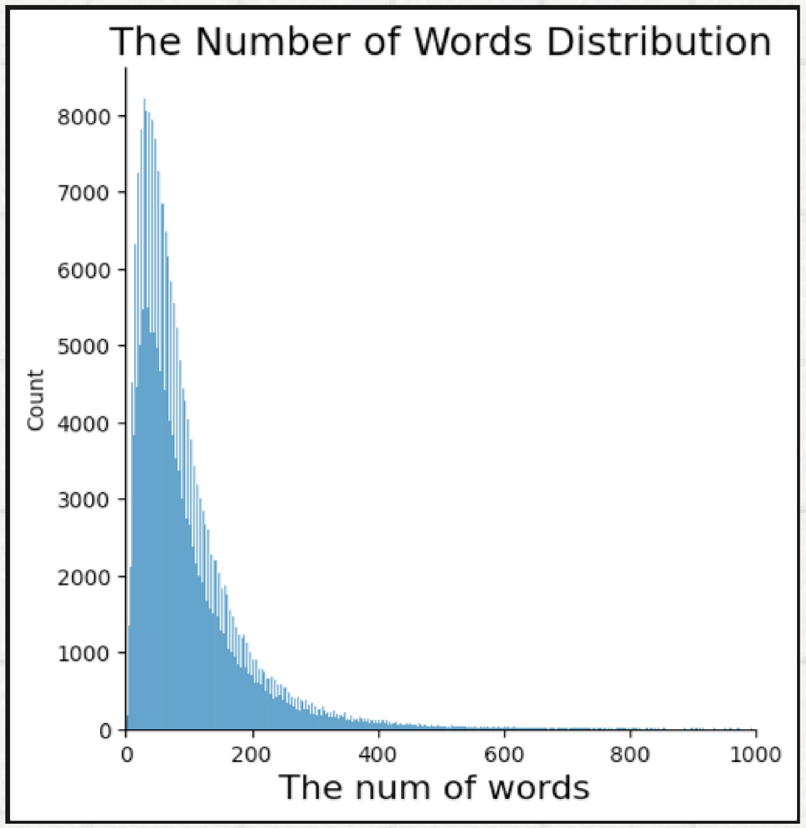
What is it?



RNN is a type of artificial neural network designed to effectively handle sequential data by considering the order and dependence of the input information.

Types of RNNs: Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs).

Recurrent Neural Network



We also ran it on Bidirectional LSTM and got our results, as we thought that a bidirectional layer in LSTM would also capture future dependecies.
But our results showed that it performed only mariginallly better than LSTM, showing that the data is much more reliant on short term dependencies more than the long term dependencies.
This affirmed our results from topic modelling which did not give us sufficiently good results as it captures long term context of group of words in the corpus.
Thus we implemented GRU which is not only computationally less expensive but cpatures short term dependencies more effectively as it is a simpler architechture.
This affirms from our data which has shorter posts in more amount.

RecurrentNeural Network

Getting a baseline

We initially ran the inbuilt LSTM and GRU models to get the baseline numbers

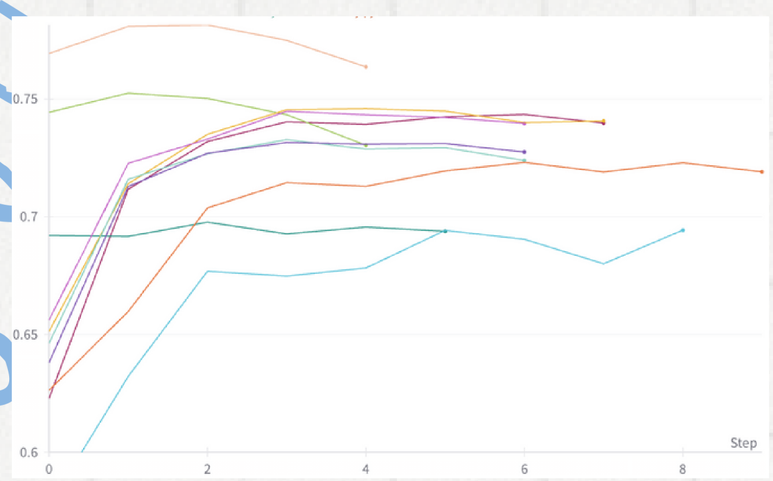
- Ran LSTM, Bidirectional LSTM and GRU to get base results

RNN		
Type of model	Accuracy	Loss
LSTM (keras)	0.732	0.7627
Bi - LSTM (keras)	0.731	0.7756
GRU (keras)	0.733	0.762

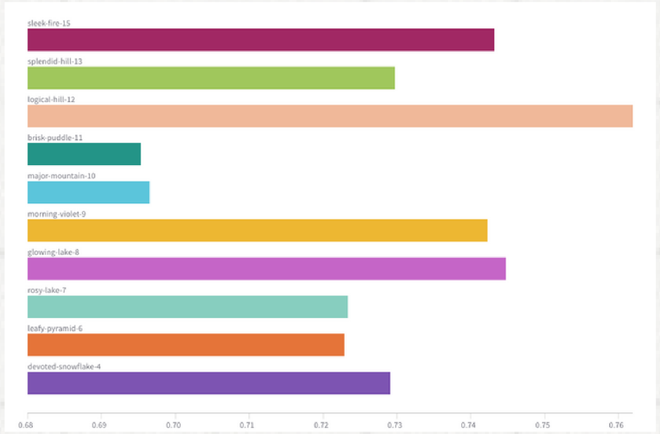
Recurrent Neural Network

Hyperparameter Tuning

Validation Accuracy



Test Accuracy



We tuned our model on the following hyperparameters:

- Units : [100, 150, **200**, 300]
- Batch size: [16, **32**,64]

Recurrent Neural Network

Building our model

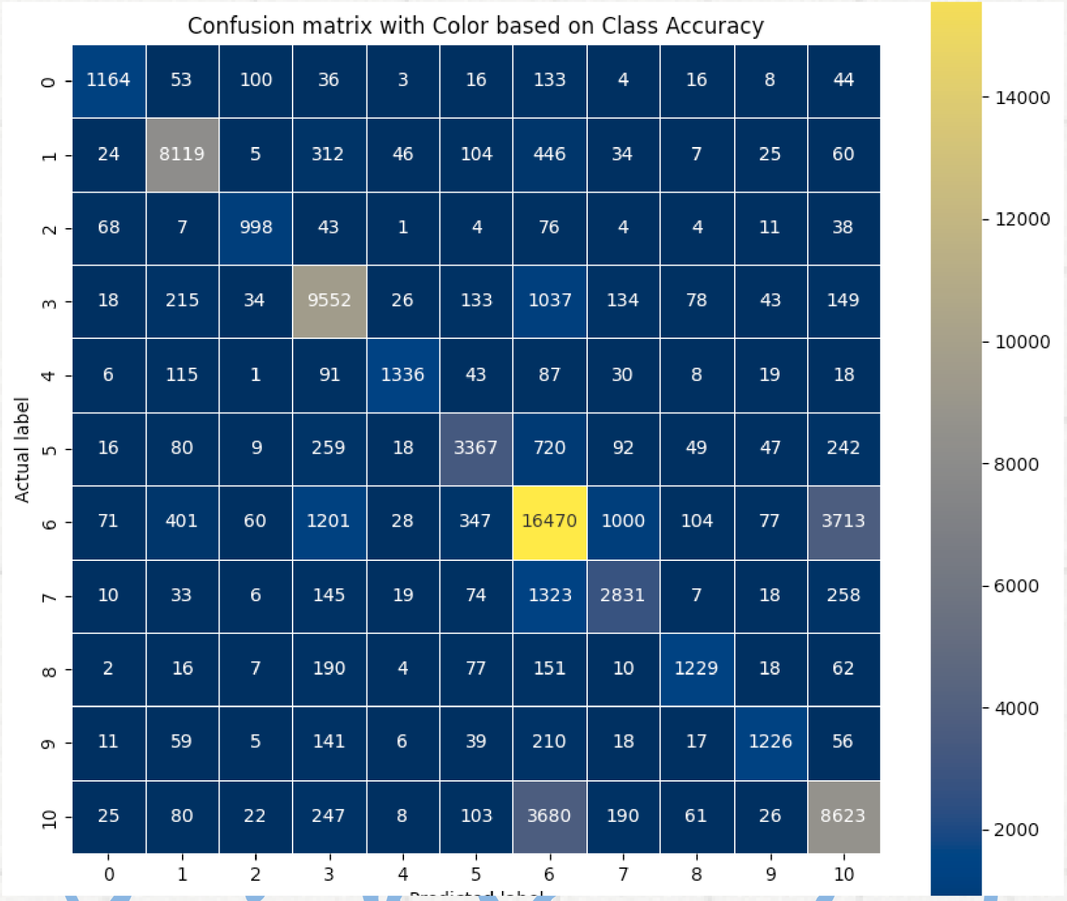
We built our model in pytorch using LSTMCell, and implemented the recurrence from scratch.
We ran it with the best hyperparameters and got the following results

RNN		
Type of model	Accuracy	Loss
LSTM (pytorch)	0.733	0.762

Recurrent Neural Network

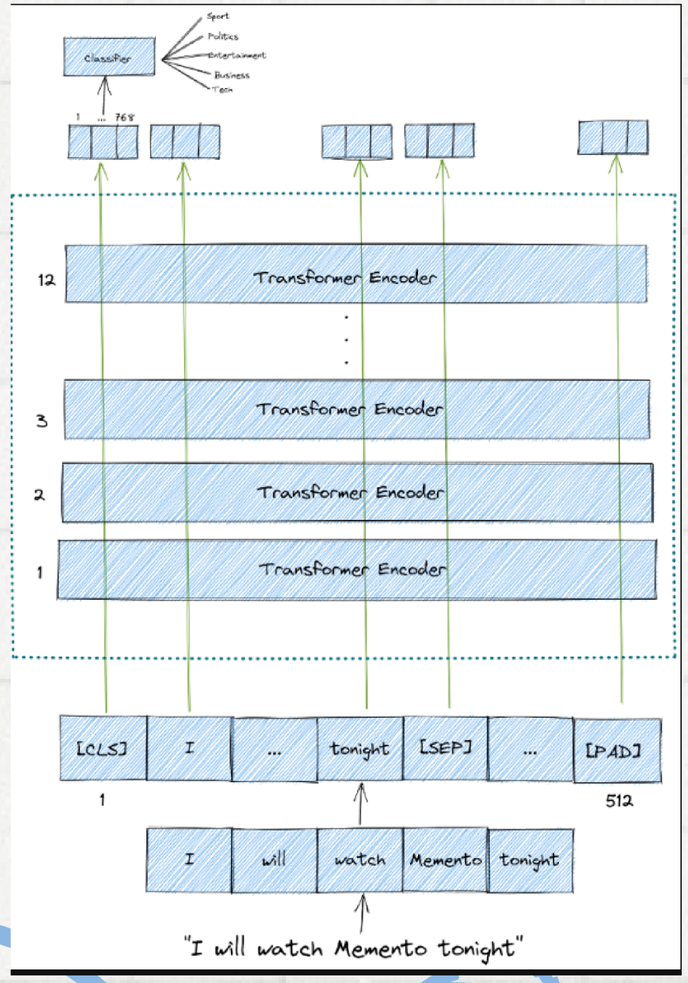
Results

RNN		
Type of model	Accuracy	Loss
LSTM (pytorch)	0.743	0.753



distilBERT

What is it?



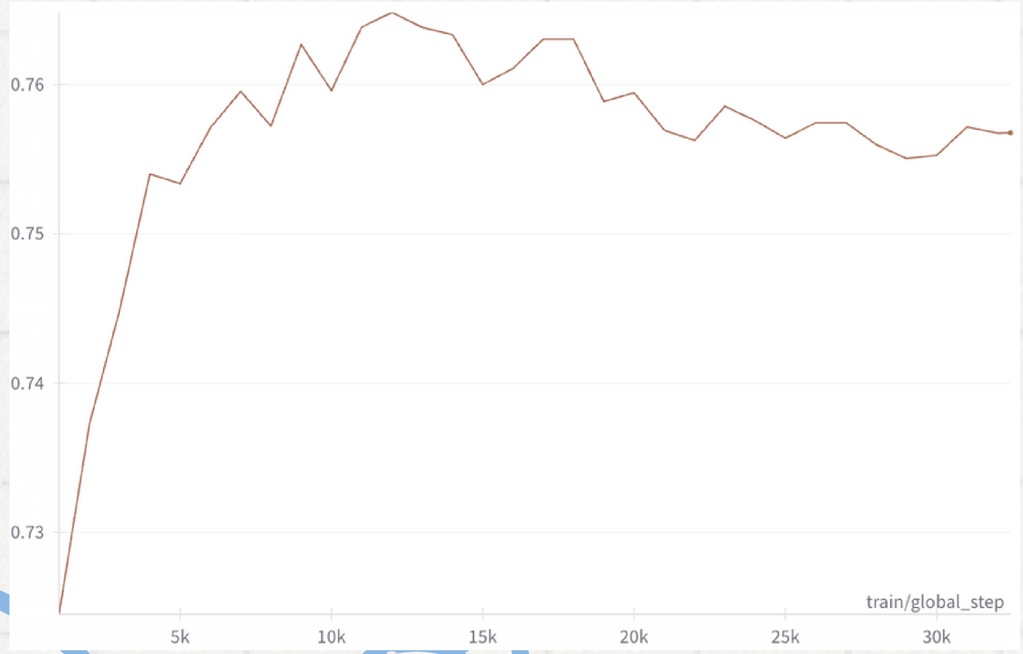
DistilBERT is a smaller and more efficient version of the BERT (Bidirectional Encoder Representations from Transformers) model, which is a state-of-the-art pre-trained language representation model. It learns contextualized representations of words or tokens by considering the entire surrounding context.

distilBERT

Initial finetuning

We finetuned the model on our data and received these results.

Validation accuracy



Validation Loss

distilBERT

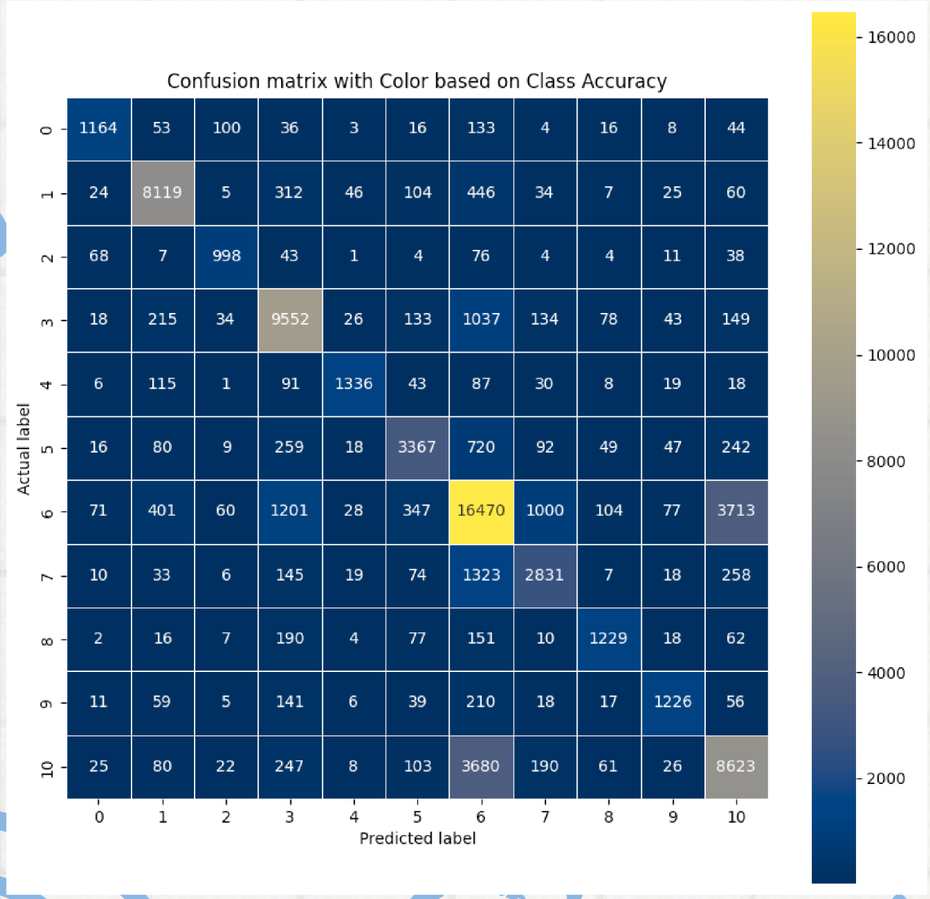
Combatting overfitting

From the results, we saw that the model was overfitting.
We decreased batch size and increased weight decay to improve regularization.

Validation accuracy

Validation Loss

Analysis of results



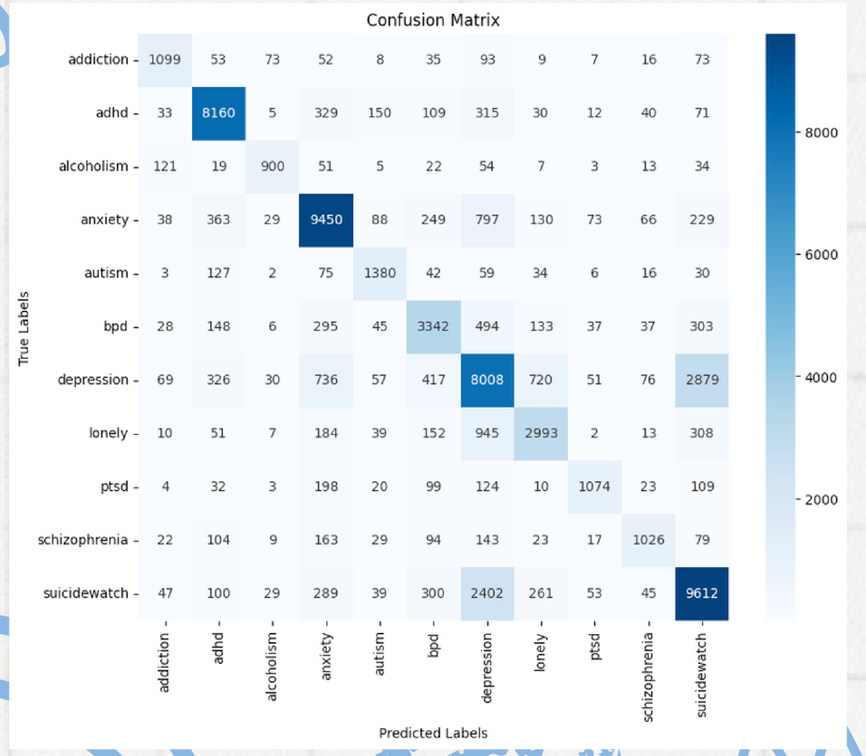
Now we see that in all of our models, there was a considerable misclassification in classes like depression and suicidewatch. Further we see that depression as it is highly represented is leading cause of misclassification in many classes.

So we decided to undersample these classes and run our models again.

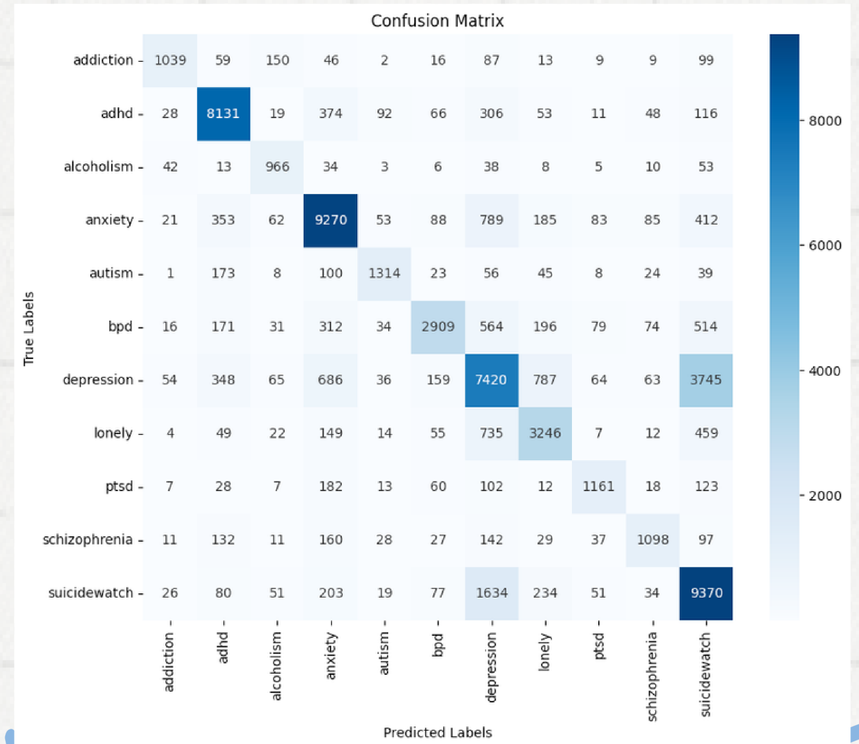
CNN with reduced data

CNN						
Type of data	Depression reduced	Suicide watch reduced	Accuracy	Precision (macro avg)	Recall (macro avg)	F-1 score (macro avg)
Data	No	No	0.668	0.68	0.67	0.67
Generated Features + Reduced Data	No	No	0.712	0.76	0.68	0.72
Generated Features + Reduced Data	Yes	No	0.725	0.75	0.70	0.72
Generated Features + Reduced Data	Yes	Yes	0.725	0.75	0.70	0.72

Only depression reduced



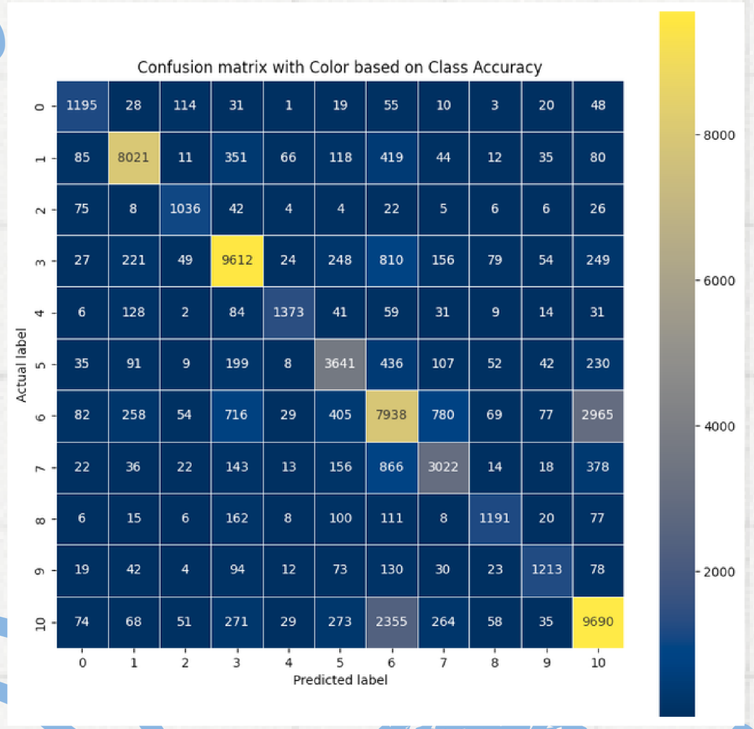
Both depression and suicidewatch reduced



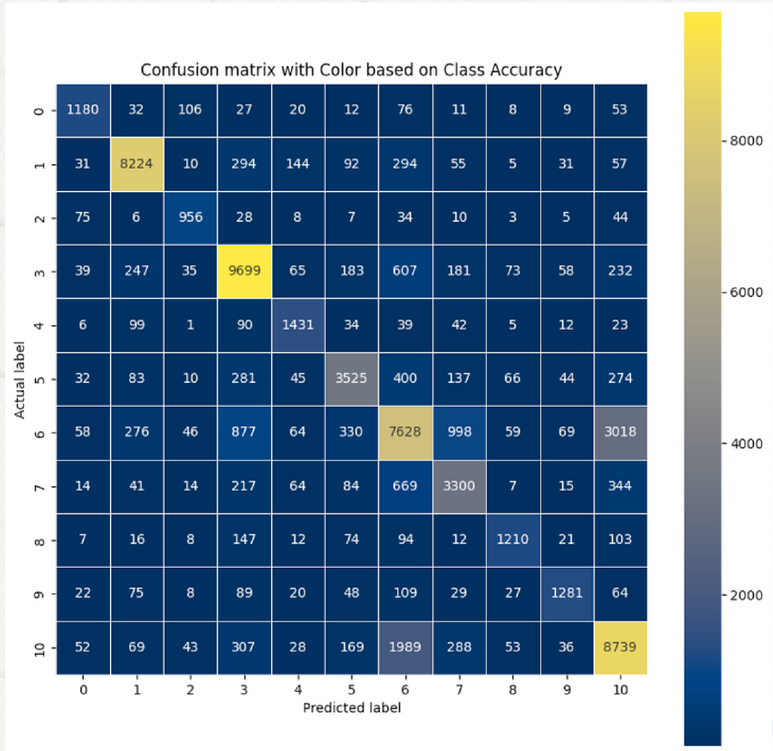
RNN with reduced data

RNN				
Type of model	‘Depression’ reduced	‘Suicidewatch’ reduced	Accuracy	Loss
LSTM (pytorch)	No	No	0.733	0.762
LSTM (pytorch)	Yes	No	0.739	0.775
LSTM (pytorch)	Yes	Yes	0.743	0.771

Only depression reduced



Both depression and suicidewatch reduced



distilBERT with reduced data

distilBERT				
Run	‘Depression’ reduced	‘Suicidewatch’ reduced	Accuracy	Loss
1	No	No	0.755	1.04
2	Yes	Yes	0.774	0.98

LIME Text Explainer

LIME (Local Interpretable Model-agnostic Explanations) is a technique used for explaining the predictions of machine learning models.
It takes in a particular post and generates 100 perturbations of the sentences. for which the model predicts outputs.
Then the model predicts for all those perturbations and lime computes probabilities of the the words which play a vital role in predicting each class and also which play a vital role in the making the model not predicting a particular class.

LIME Text Explainer

Post 1

Every Waking Moment I know whether proper protocol I new Reddit really know way around set rule I missed I broken one I apologise But since I waded started replying I thought polite introduce Hi I Strix I woman living UK I experiencing suicidal ideation year lived depression probably life Over last month particularly thought become much pronounced point shading every waking moment What idle notion became expectation without particular intent last week solidified I describe desire I know causing I lucky life I stable job I experienced abuse violence life I traumatic event I recall I married I love wife much I lucky loved I kid I dear little distant extended family particular money concern beyond usual household So come total mystery How got stage I know I fairly open I think I made decision tell wife I tried balance whether better knowing taken completely surprise anything happened I still sure I right thing I know I She supportive employer also aware made adjustment I seen doctor counsellor I literally nothing complain I compulsion This expectation And recently developing intention I keep asking real If I sure I really feeling I think I I fishing attention After I talk freely real life I still come group like What I looking Practical advice I think I everything practical advice would probably suggest But still I find I timescale mind month away yet many yearly event coming soon I want taint people I idea method I looking around suitable place But I never attempted anything yet Too frightened Still frightened So I know Is even real Or I mood When time come I still want For I seem able think anything else It feel scary little bleak guilty I mind admitting But along something else Something like anticipation The thought fulfilling something It weird But time I I hope nobody mind I pop group called A group A sub Have learned lingo yet occasionally I ask let know I anything wrong including post Thanks reading wall text

Actual label - Suicidewatch

LIME Text Explainer

RNN

Predicted label - Depression

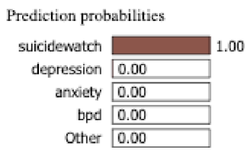
Text with highlighted words

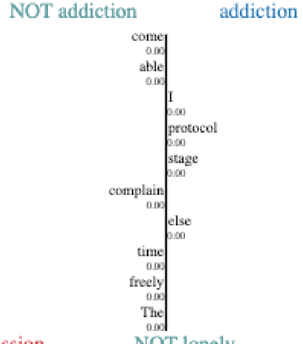
? ? ? ? ? ? ? ? ? every waking moment dont know whether proper protocol im new reddit dont really know way around theres set rules ive missed ive broken one apologisebut since ive waded started replying thought itd polite introduce hi im stri im woman midforties living uk ive eperiencing suicidal ideation year lived depression probably 3 4 life last months particularly thoughts become much pronounced point theyre shading every waking moment idle notion became epectation without particular intent last weeks solidified describe desirei dont know whats causing ive lucky life stable job ive eperienced abuse violence life ive traumatic events recall im married love wife much im lucky loved kids dear little distant etended family particular money concerns beyond usual gettingby householdso comes total mystery got stage dont know ive fairly open think made decision tell wife tried balance whether shed better knowing taken completely surprise anything happened im still sure right thing know shes supportive employers also aware made adjustments ive seen doctors counsellorsi literally nothing complain compulsion epectation recently developing intentioni keep asking real im sure im really feeling think im fishing attention talk

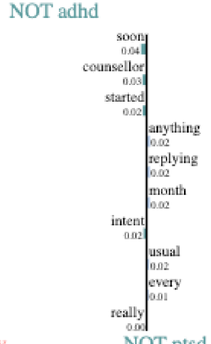
LIME Text Explainer

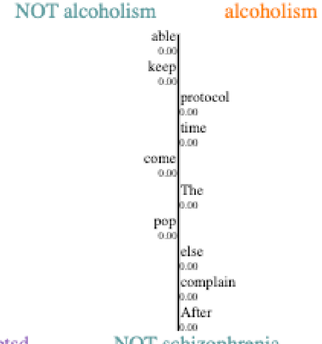
distilBERT

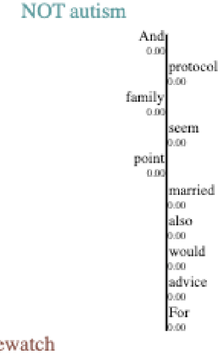
Predicted label - Suicidewatch

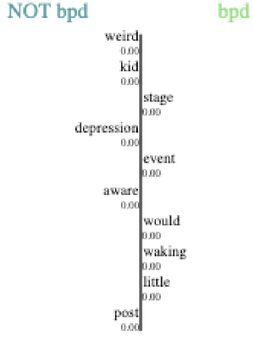


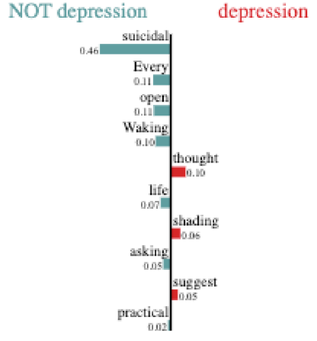


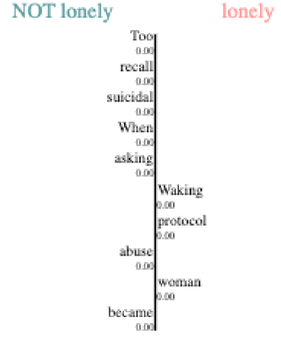


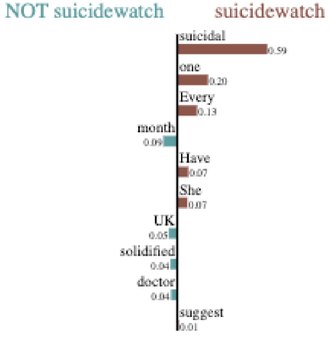












Text with highlighted words

Every Waking Moment I know whether proper protocol I new Reddit really know way around set rule I missed I broken one I apologise But since I waded started replying I thought polite introduce Hi I Strix I woman living UK I experiencing suicidal ideation year lived depression probably life Over last month particularly thought become much pronounced point shading every waking moment What idle notion became expectation without particular intent last week solidified I describe desire I know causing I lucky life I stable job I experienced abuse violence life I traumatic event I recall I married I love wife much I lucky loved I kid I dear little distant extended family particular money concern beyond usual household So come total mystery How got stage I know I fairly open I think I made decision tell wife I tried balance whether better knowing taken completely surprise anything happened I still sure I right thing I know I She supportive employer also aware made adjustment I seen doctor counsellor I literally nothing complain I compulsion This expectation And recently developing intention I keep asking real If I sure I really feeling I think I I fishing attention After I talk freely real life I still come group like What I looking Practical

LIME Text Explainer

Post 2

feeling sometimes feel wan na die like actually wan na die want everything stop sleep deprived barely see sleep go boring life heavy head heavy feeling suddenly think start realize YOUR action decision made turn like bc best pray god soon something good happens life let know right thing wish told one person much mean still time wish spent time loving spending time still could got late

Actual label - Depression

LIME Text Explainer

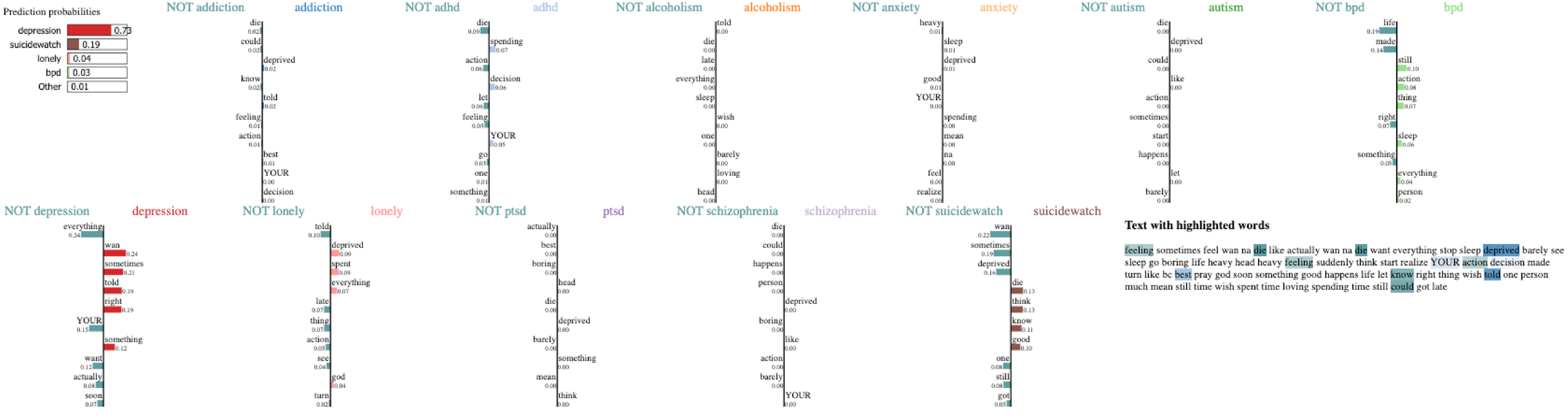
RNN

Predicted label – Depression

LIME Text Explainer

distilBERT

Predicted label - Depression



Stress Testing

Motivation

We also tested the robustness of our best model via stress testing, which included synonym replacement and masking.
Synonym replacement included randomly replacing 10% OF THE data with its synonym.
As in many post we saw that they have direct label name contained in them, which makes it easier for the model to learn from and classify.
Hence we masked these words with a 75% chance.

Stress Testing

Results

Results of stress testing

```
base_acc = compute_accuracy(test_labels)

print("Base accuracy: ", base_acc)
```

```
Base accuracy:  0.8628170894526035
```

```
syn_acc =  compute_accuracy(test_labels , test='syn')

print('Synonym Accuracy: ', syn_acc)
```

```
Synonym Accuracy:  0.8110814419225634
```

```
mask_acc =  compute_accuracy(test_labels, test='mask')

print('Mask Accuracy: ', mask_acc)
```

```
Mask Accuracy:  0.815086782376502
```

Final reflections and future steps

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.