Distinguishing Mental Health Categories on Reddit Using NLP

Spoorthi Jayaprakash Malgund (SUID: 904543696) - Team Lead
Aditee Suryakant Malviya (SUID: 465609468)
Adithya Neelakantan (SUID: 663182945)
Krupa Minesh Shah (SUID: 619191197)
Raj Nandini (SUID: 949041747)
Harshitha Reddy Kancherla (SUID: 519904638)
Mrunmayee Atul Jakate (SUID: 652397247)

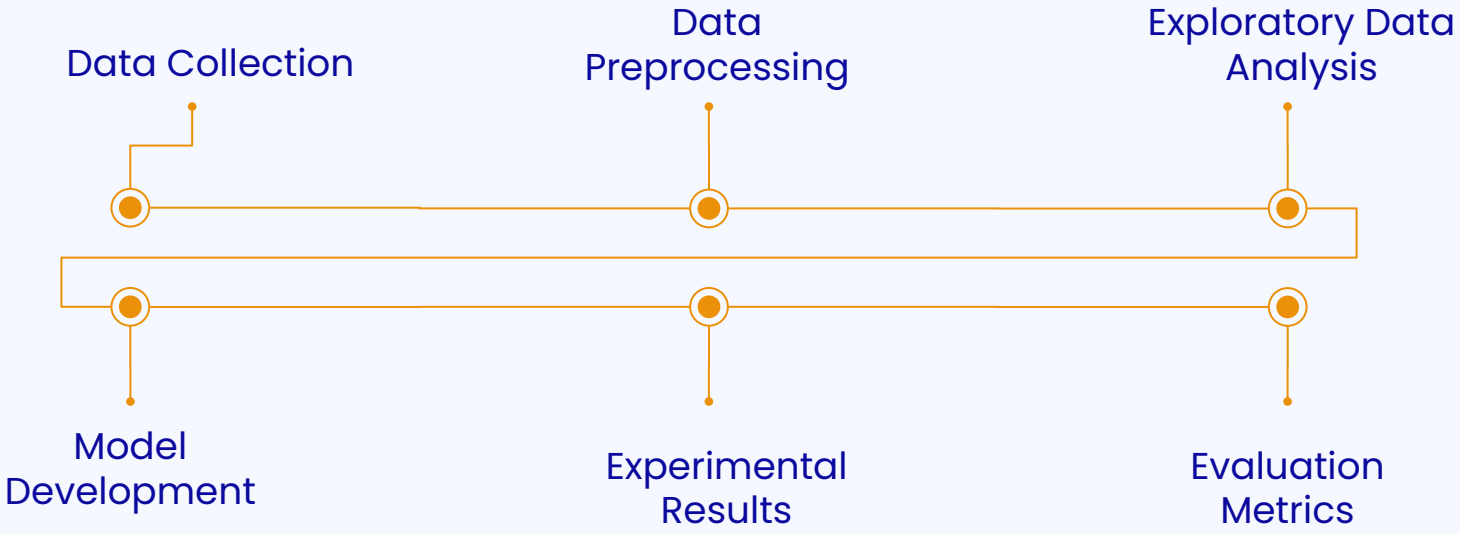
Introduction

- An opportunity to understand how people express challenges related to conditions like depression, anxiety, and suicidal ideation.
- Aimed to develop a model that can better classify and analyze specific mental health conditions based on the conversations being had.
- Subreddits used : **['BPD', 'SuicideWatch', 'depression', 'Anxiety', 'CasualConversation', 'mentalhealth', 'bipolar']**
- Not only improve understanding of online mental health dynamics but may also support intervention efforts, foster more supportive and safe digital spaces for those seeking help.

Problem Statement

- Classify posts into specific mental health conditions: anxiety, depression and healthy.
- Assess severity levels, including identifying high-risk indicators like suicidal ideation.
- Aims to support researchers and mental health professionals in better understanding and responding to mental health discussions online.

Methodology



Data Collection

	title	body	author	url	score	upvote_ratio	subreddit
0	Craving closeness	About a year ago, I got into a relationship th...	Nash4N00b	https://www.reddit.com/r/mentalhealth/comments...	2	1.00	mentalhealth
1	Mental health class	I think schools should have a mandatory mental...	ConnecticutJohn	https://www.reddit.com/r/mentalhealth/comments...	112	0.97	mentalhealth
2	My 7yo son accidentally fell on our pet bird a...	It was TRULY a freak accident and not expected...	AnonymousCounselor52	https://www.reddit.com/r/mentalhealth/comments...	387	0.97	mentalhealth
3	Who else doesn't want covid world to leave	My mental health has been so bad and I've been...	jjdabest8820	https://www.reddit.com/r/mentalhealth/comments...	15	0.94	mentalhealth
4	my mental health is suffering due to my contro...	note: please read my previous posts for contex...	Wooden_View_7120	https://www.reddit.com/r/mentalhealth/comments...	7	0.83	mentalhealth

- Used Reddit's PRAW API, data gathered from mental health-focused subreddits (e.g., r/depression, r/Anxiety, r/SuicideWatch) and general discussions (e.g., r/CasualConversation) for balanced control data.
- Subreddits serve as labels – to represent specific mental health conditions, severity levels, or healthy/non-healthy classifications.

Data Preprocessing

Handling Missing Data	Removed rows with missing or irrelevant content, especially in the body field. For subreddits like bipolar and mentalhealth, excluded posts with missing body text.
Data Cleaning	Removed Reddit-specific formatting, such as markdowns, URLs, and special characters. Tokenized text by splitting it into individual words. Eliminated stopwords (e.g., "and," "the") and punctuation to reduce noise.
Lemmatization	Reduced words to their base forms to retain core meaning. Focused on key parts of speech (nouns and adjectives) to emphasize meaningful content.
Data Partitioning	Split the dataset into training (80%), validation (10%), and testing (10%) subsets. Ensured sufficient representation of each label in all partitions.
Final Dataset	Post-Cleaning Shape: Approximately 6,000 posts after cleaning and preprocessing. Balanced and structured for effective model training and evaluation.

Data Preprocessing

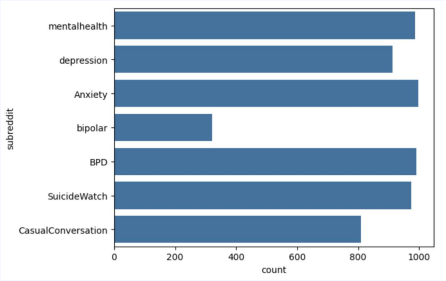
Lemmatization

url	full_text	step_01_reddit_cleanup	step_02_remove_stopwords	step_03_remove_punctuation	lemmatization_standard	lemmatization_noun_adj
https://www.reddit.com/r/mentalhealth/comments...	Craving closeness About a year ago, I got into...	Craving closeness About a year ago, I got into...	Craving closeness About year ago, I got relati...	Craving closeness About year ago I got relatio...	crave closeness about year ago I get relations...	closeness year relationship future tough decis...
https://www.reddit.com/r/mentalhealth/comments...	Mental health class I think schools should hav...	Mental health class I think schools should hav...	Mental health class I think schools mandatory ...	Mental health class I think schools mandatory ...	mental health class I think school mandatory m...	mental health class school mandatory mental he...
https://www.reddit.com/r/mentalhealth/comments...	My 7yo son accidentally fell on our pet bird a...	My 7yo son accidentally fell on our pet bird a...	My 7yo son accidentally fell pet bird killed 1...	My 7yo son accidentally fell pet bird killed 1...	my 7yo son accidentally fall pet bird kill 11 ...	7yo son bird son help advice trauma first expe...
https://www.reddit.com/r/mentalhealth/comments...	Who else doesn't want covid world to leave My ...	Who else doesn't want covid world to leave My ...	Who else doesn't want covid world leave My men...	Who else doesn't want covid world leave My men...	who else do not want covid world leave my ment...	covid world mental health bad battle life anxi...
https://www.reddit.com/r/mentalhealth/comments...	my mental health is suffering due to my contro...	my mental health is suffering due to my contro...	mental health suffering due controlling toxic ...	mental health suffering due controlling toxic ...	mental health suffering due control toxic hous...	mental health suffering due toxic household no...

- Reduced words to their base forms to retain core meaning.
- Focused on key parts of speech (nouns and adjectives) to emphasize meaningful content.

Exploratory Data Analysis (EDA)

Visualizations

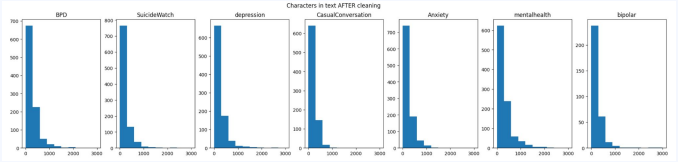


1. Subreddit Distribution :

Bar chart for visualization highlighting class imbalance.
Post-cleaning adjustments led to fewer posts overall but preserved class diversity.

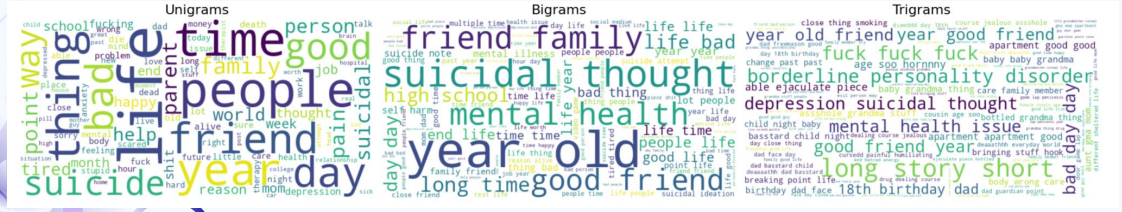
2. Text Length Analysis:

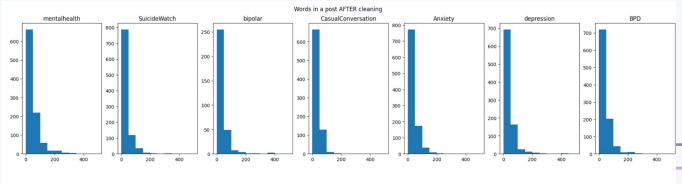
Histograms of word and character counts before and after cleaning.



3. Word Clouds:

Unigrams: Common terms like "life," "suicide," "depression".
Bigrams: Frequent phrases such as "mental health," "need help."
Trigrams: Examples include "depression suicidal thought," showing contextual information.





Exploratory Data Analysis (EDA)

Statistical Insights

1. **TF-IDF Analysis:**
   Measured term importance within each subreddit by calculating Term Frequency-Inverse Document Frequency (TF-IDF) scores.

- r/SuicideWatch: High TF-IDF words like "life," "people," "suicide," reflecting crisis themes.
- r/depression: High TF-IDF words like "life," "time," "depression," showing emotional struggles.

Low TF-IDF words in both subreddits were rare or irrelevant terms.

2. **Sentiment Analysis**

Measured average sentiment scores across subreddits:
- r/CasualConversation: Positive sentiment.
- r/SuicideWatch and r/depression: Negative sentiment.

subreddit	
Anxiety	−0.300415
BPD	−0.165735
CasualConversation	0.464521
SuicideWatch	−0.456508
bipolar	−0.122982
depression	−0.331703
mentalhealth	−0.218340

3. **Vocabulary Diversity**:
- Unique word counts for subreddits showed high lexical variety across all classes.

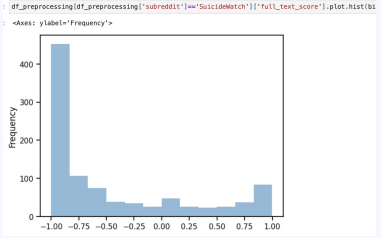
---- depression ----

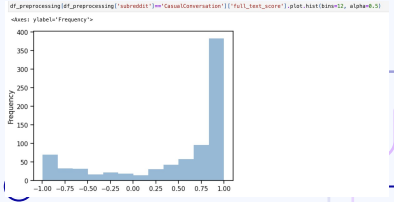
Words with highest TF-IDF score

	word	tf	idf	tfidf
0	life	782	1.866093	41.811753
1	time	673	1.944618	35.104992
2	people	539	2.171289	32.503732
3	day	521	2.181946	31.076752
4	thing	550	2.133156	30.213537
5	friend	485	2.263776	29.749695
6	depression	359	2.307348	27.727715
7	year	469	2.225749	26.635652
8	bad	359	2.369999	26.053195
9	good	351	2.400635	23.884599

Words with lowest TF-IDF score

	word	tf	idf	tfidf
4492	sand	1	7.123589	0.073119
4493	scatter	1	7.123589	0.073119
4494	sighing	1	7.123589	0.073119
4495	slim	1	7.123589	0.073119
4496	straightup	1	7.123589	0.073119
4497	twat	1	7.123589	0.073119
4498	typing	1	7.123589	0.073119
4499	wasteland	1	7.123589	0.073119
4500	watercolor	1	7.123589	0.073119
4501	zombielike	1	7.123589	0.073119



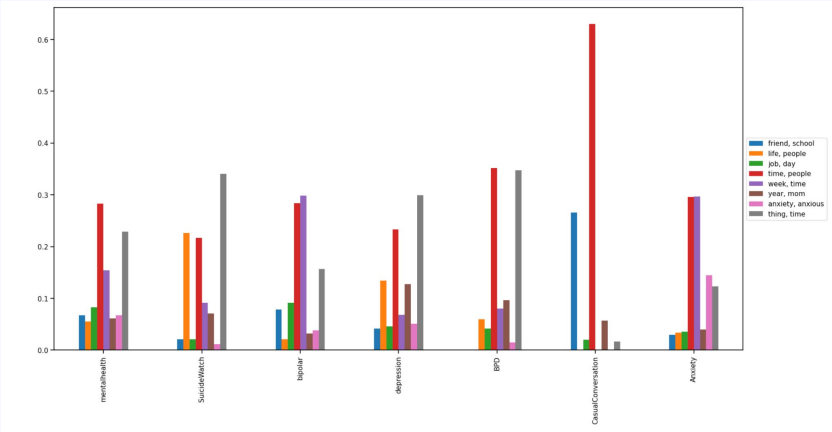


Exploratory Data Analysis (EDA)

Topic Modeling

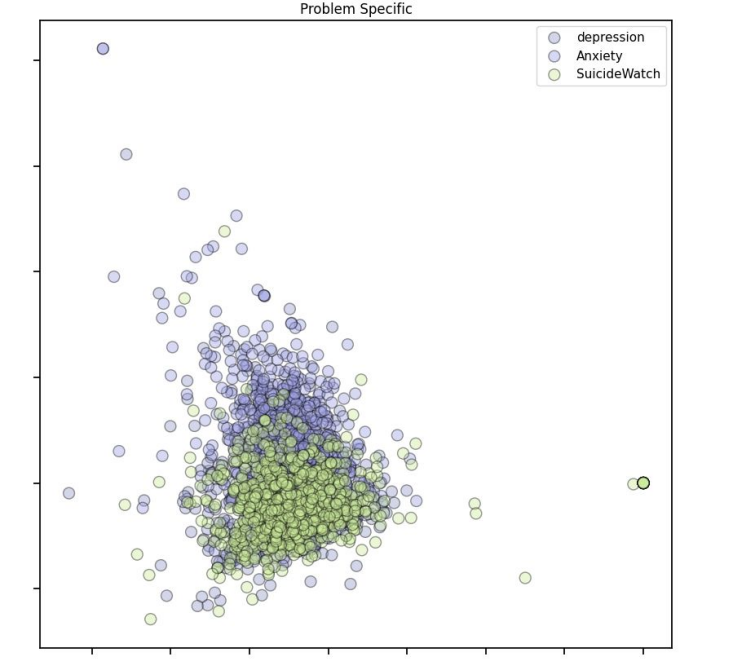
Author Topic Modeling

- Author-Topic Modeling is a technique that analyzes the distribution of topics across documents, considering the authorship of the text.
- Highlights dominant themes, such as general conversations across all subreddits.
- Therapy and medication discussions are prevalent in r/bipolar but absent in r/SuicideWatch and minimal in r/depression, reflecting hopelessness. In r/Anxiety, users frequently seek help by exposing their problems, while r/CasualConversation contains mostly positive discussions.



Exploratory Data Analysis (EDA)

Dimensionality Reduction



- Dimensionality reduction simplifies the dataset by converting it into a lower number of features (dimensions) while keeping as much meaningful information as possible.
- Reddit Post dimensionality reduction of the data it was evident that differentiating between "depression" and "suicide" is quite challenging, likely because the latter is inherently a subset of the former.
- Noisy labels make this relationship unclear, as seen in this sample plot.
- To bypass this challenge, we combine the r/depression and r/SuicideWatch under the common label “depression” for the baseline models.

Model Development

Comparative Analysis

- **Two baselines – Classification using Reduced Labels ('r/depression', 'r/Anxiety', 'r/CasualConversation')**
  1. Most frequent Label Classifier
  2. Logistic Regression with 2-6 gram character TF-IDF features

- **Advanced Models – Structured Predictions**
  1. Structured Perceptron
  2. LSTM (Long Short-Term Memory)

Model Development

Baseline Model with Reduced Labels

Objective

Simplify the classification task by grouping r/depression and r/SuicideWatch under a common label, "depression," resulting in three labels: Anxiety, Depression, and Normal.

Baseline Models

1. **Most Frequent Label Classifier:**
    - Assigns the most common label (depression) to all predictions.
    - Uses label frequency as a naive baseline for comparison.
2. **Logistic Regression with TF-IDF:**
    - **TF-IDF Vectorization:**
        - Converts text data into numerical features using term importance.
        - Applied at:
            - Post Level: Focuses on individual terms and phrases for precise representation.
            - Document Level: Captures contextual relationships across aggregated text.
    - **Logistic Regression with extracted TF-IDF features:**
        - Effective in capturing direct correlations between terms and labels but limited in handling complex or sequential patterns.

Logistic Regression with TF-IDF offers a strong baseline but lacks the ability to capture complex relationships and sequential patterns, prompting the use of advanced models like Perceptron and LSTM.

Model Development

Advanced Models

Structured Perceptron

- **Categorizing into Labels LDA (Latent Dirichlet Allocation):**
  - A: Mental health-related, B: not related to mental health, O: Neutral or ambiguous.
  - LDA enables automatic generation of labeled sequences without manual annotation, crucial for structured prediction.

  Example Sentences – "I feel very anxious today." , "Let's talk about the weekend."
  Pseudo Tags – "I feel very A today." , "Let's talk about the B."
  Sequences – ["O", "O", "O", "A", "O"], [["O", "O", "O", "O", "B"]

- **Structured Perceptron**
  - Considers context within sentence sequences.
  - Feature functions are combined with weights that get iteratively updated to compute scores.
  - Updates using Viterbi decoding to find the best sequences. Correct sequence increases score.
  - Model predicts best label sequence by maximizing score.

  Captures dependencies between words in a sentence, unlike independent classifiers.
  Ensures robust handling of sequential data, especially for nuanced mental health posts.

Model Development

Advanced Models

LSTM (Long Short-Term Memory)

- Capture long-term dependencies in text, ideal for mental health posts with sequential data.
- Input sequences labeled (A, B, O) and padded to a maximum length identified during data exploration.
- Embedding layer to convert words into dense vector representations (Predicts whether a word (tags) is likely to appear nearby).
- Dense output layer for tag prediction (A,B,O) for each position in the sequence.
- Training Loss Function – Categorical cross-entropy across multi-class tags.
- Optimizes weights through backpropagation through time (BPTT).

Better than the baseline models in sequential categorical token data, especially for long reddit posts where there might be a longer update that depends on contextual and temporal dependencies and clues in token sequences. (Mental health worsening over time for example).

Maintains information over long sequences without vanishing gradients.

Evaluation Metrics

```
baseline_predictions = baseline_majority_classifier(X_train_reduced, X_dev_reduced, y_train_reduced, y_dev_reduced)

                   precision    recall  f1-score   support

          Anxiety       0.00      0.00      0.00       204
CasualConversation       0.00      0.00      0.00       175
       depression       0.49      1.00      0.66       360

         accuracy                           0.49       739
        macro avg       0.16      0.33      0.22       739
     weighted avg       0.24      0.49      0.32       739

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1531: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

Logistic Regression with TF-IDF at post-level

```
X_train_reduced_post, X_dev_reduced_post, X_test_reduced_post = char_ngrams_tfidf_vectorizer(X_train_reduced, X_dev_reduced, X_test_reduced)
```

```
lr_tfidf_post_predictions = train_and_evaluate(X_train_reduced_post, X_dev_reduced_post, y_train_reduced, y_dev_reduced, custom_model=None, return_predictions=True)

/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:1247: FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed in 1.7. From then on, it will always use 'multinomial'. Leave it to its default value to avoid
  warnings.warn(
                   precision    recall  f1-score   support

          Anxiety       0.91      0.73      0.81       204
CasualConversation       0.75      0.52      0.61       175
       depression       0.73      0.93      0.82       360

         accuracy                           0.78       739
        macro avg       0.80      0.72      0.75       739
     weighted avg       0.79      0.78      0.77       739
```

Logistic Regression with TF-IDF at document-level

```
X_train_reduced_docs, X_dev_reduced_docs, X_test_reduced_docs = vectorize_documents_by_subreddit(X_train_reduced, y_train_reduced, X_dev_reduced, X_test_reduced)
```

```
train_and_evaluate(X_train_reduced_docs, X_dev_reduced_docs, y_train_reduced, y_dev_reduced, custom_model=None)

/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:1247: FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed in 1.7. From then on, it will always use 'multinomial'. Leave it to its default value to avoid
  warnings.warn(
                   precision    recall  f1-score   support

          Anxiety       0.94      0.39      0.55       204
CasualConversation       0.50      0.01      0.01       175
       depression       0.55      0.99      0.71       360

         accuracy                           0.59       739
        macro avg       0.66      0.46      0.42       739
     weighted avg       0.65      0.59      0.50       739
```

Evaluation Metrics

```
sp = StructuredPerceptron()
inference_method = 'greedy'
%time sp.fit(instances_train, instances_test, iterations=10, inference=inference_method)

sp.save('model_greedy.pickle')
```

```
.................................................1000
.................................................2000
.................................................3000
.........................   314965 features
        Training accuracy: 0.91

        Development accuracy: 0.88

.................................................1000
.................................................2000
.................................................3000
.........................   320913 features
        Training accuracy: 0.92

        Development accuracy: 0.89

.................................................1000
.................................................2000
.................................................3000
.........................   323672 features
        Training accuracy: 0.92

        Development accuracy: 0.88

.................................................1000
.................................................2000
.................................................3000
.........................   325370 features
        Training accuracy: 0.93

        Development accuracy: 0.88

.................................................1000
.................................................2000
.................................................3000
.........................   326442 features
        Training accuracy: 0.93

        Development accuracy: 0.86

.................................................1000
.................................................2000
.................................................3000
.........................   327104 features
        Training accuracy: 0.94
```

```
[ ] print(sp.predict('My life is so amazing!'.split(), method='greedy'))
['0', 'B', '0', '0', '0']

[ ] print(sp.predict('The medication is not working'.split(), method='greedy'))
['0', '0', '0', 'A', '0']
```

```
Epoch 1/5
180/180 ━━━━━━━━━━━━━━━━━━━━ 406s 2s/step - accuracy: 0.0576 - loss: 0.6735 - val_accuracy: 0.0589 - val_loss: 0.1661
Epoch 2/5
180/180 ━━━━━━━━━━━━━━━━━━━━ 431s 2s/step - accuracy: 0.0577 - loss: 0.1557 - val_accuracy: 0.0584 - val_loss: 0.1581
Epoch 3/5
180/180 ━━━━━━━━━━━━━━━━━━━━ 392s 2s/step - accuracy: 0.0594 - loss: 0.1413 - val_accuracy: 0.0587 - val_loss: 0.1522
Epoch 4/5
180/180 ━━━━━━━━━━━━━━━━━━━━ 445s 2s/step - accuracy: 0.0583 - loss: 0.1404 - val_accuracy: 0.0588 - val_loss: 0.1546
Epoch 5/5
180/180 ━━━━━━━━━━━━━━━━━━━━ 438s 2s/step - accuracy: 0.0589 - loss: 0.1315 - val_accuracy: 0.0601 - val_loss: 0.1547
```

```
Enter sentences separated by commas: i, am, sad
[(['i'], ['0']), (['am'], ['0']), (['sad'], ['A'])]
```

Conclusion

1. Structured Perceptron and LSTM captured contextual and temporal dependencies, **outperforming baseline models** in handling sequential data.

2. LSTM models **excel at maintaining context across longer sequences**, which is critical for Reddit posts where mental health themes may develop over time.

3. LSTM captures temporal dependencies between words, ensuring it understands how **earlier tokens influence later ones**, crucial for nuanced topics like mental health.

Future Scope

- **Collaboration Opportunities**: Partner with mental health professionals to integrate domain expertise into model development.

- **Advanced Models**: Explore deep learning architectures like transformers for better understanding of complex language patterns and long-term dependencies.

- **Real-World Application**: Integrate the model into mental health platforms for early intervention, enabling practical support and research in mental health trends.