Harnessing NLP to Detect Stress

in

Social Media

Early Intervention for Mental Wellbeing

Chong Chee Chen (Jimmy)

19 Oct 2024

Table of Contents

1. Problem Statement

This project investigates the feasibility of automatically detecting stress in social media text using Natural Language Processing (NLP) techniques. Early detection of stress can be crucial for improving mental health outcomes. Social media platforms offer a vast amount of user-generated text data that can be potentially analysed to identify individuals who might be experiencing stress.

Currently, identifying stress in individuals often relies on self-reporting or clinical assessments. These methods can be subjective and time-consuming. Analysing social media data offers a potential avenue for more objective and scalable stress detection.

The desired state is to develop a reliable and efficient NLP-based system that can automatically detect stress in social media text. This system could be used to:

- Flag individuals who might be experiencing stress for further evaluation or intervention.
- Provide targeted support resources to those in need.
- Gain insights into population-level stress trends.

There has been some research on stress detection using social media data. Existing studies have explored various NLP techniques, such as sentiment analysis. This project aims to build upon this existing work by exploring the use of more advanced techniques like TF-IDF vectorization and machine learning models for improved accuracy.

2. Industry & Domain

The project primarily falls within the **Natural Language Processing (NLP)** and **Mental Health** domains. Specifically, it focuses on applying NLP techniques to social media data to detect stress.

The NLP industry is rapidly evolving, driven by advancements in machine learning and deep learning algorithms. While there have been significant breakthroughs, challenges remain in areas such as handling ambiguity, context understanding, and domain adaptation.

- **Mental Health:** The mental health landscape is facing increasing pressure due to rising rates of mental health disorders and the stigma associated with seeking help. There is a growing need for innovative solutions that can provide accessible and effective mental health support.

2.1. Value Chain

- **Data Collection:** Gathering social media data from platforms like Reddit and Twitter.
- **Data Preprocessing:** Cleaning, filtering, and preparing the data for analysis.
- **Feature Engineering:** Creating relevant features from the text data (e.g., TF-IDF vectors, word embeddings).
- **Model Development:** Training and evaluating NLP models (e.g., LinearSVC, LSTM) to classify text as indicative of stress or not.
- **Deployment:** Integrating the trained model into a practical application, such as a chatbot or a web-based tool.
- **Evaluation and Refinement:** Continuously monitoring the model's performance and making improvements based on feedback.

2.2. Key Concepts

- **NLP:** Tokenization, stemming, lemmatization, stop word removal, feature extraction (TF-IDF, word embeddings), machine learning algorithms (e.g., SVM, Naive Bayes, neural networks).
- **Mental Health:** Stress, anxiety, depression, emotional well-being, mental health resources.

2.3. Potential Applications

- **Customer Service:** Analysing customer feedback to identify areas of frustration or dissatisfaction.
- **Human Resources:** Monitoring employee sentiment and identifying potential signs of burnout.

- **Healthcare:** Analysing patient records to detect early signs of mental health conditions.
- **Marketing:** Understanding consumer sentiment towards products or brands.

3. Stakeholders

- **Individuals:**
  - Social media users who may be experiencing stress.
  - Mental health professionals and organisations.
  - Family members and friends of individuals struggling with mental health.
- **Organisations:**
  - Social media platforms (e.g., Reddit, Twitter).
  - Mental health research institutions.
  - Technology companies developing AI-powered mental health solutions.
  - Healthcare providers and insurers.
  - Government agencies responsible for mental health initiatives.

3.1. Stakeholder Concerns and Motivations

- **Individuals:**
  - Early detection of stress can lead to timely intervention and improved mental health outcomes.
  - Access to personalised support resources can help individuals cope with stress and prevent it from escalating into more severe conditions.
- **Mental health professionals and organisations:**
  - This technology can supplement traditional mental health assessments and provide a more objective and scalable approach to identifying individuals at risk.
  - Early intervention can reduce the burden on mental health services and improve overall outcomes.
- **Organisations:**
  - Social media platforms have a responsibility to promote the well-being of their users.
  - Technology companies can leverage their expertise to develop innovative solutions for mental health challenges.
  - Healthcare providers and insurers can benefit from tools that can help identify individuals who may require mental health services.
  - Government agencies can use this technology to inform public health policies and allocate resources effectively.

3.2. Stakeholders’ Expectations

- **Individuals:**
  - Accuracy and reliability of the stress detection system.
  - Privacy and confidentiality of personal data.
  - Accessibility and ease of use of the technology.
- **Mental health professionals and organisations:**
  - Integration with existing mental health services.
  - Evidence-based approach to the development and validation of the technology.
  - Ethical considerations regarding the use of AI in mental health.
- **Organisations:**
  - Scalability and cost-effectiveness of the solution.
  - Positive impact on user engagement and well-being for social media platforms.
  - Contribution to advancements in AI and mental health research.
  - Alignment with ethical guidelines and regulations related to data privacy and mental health.

4. Business question

4.1. Core Business Question

Can we accurately detect stress in social media text using NLP techniques, and if so, how can this information be used to improve mental health outcomes?

4.2. Quantifying Business Value

- **Social media platforms:** By identifying users who may be experiencing stress, platforms can provide targeted support resources and potentially reduce negative outcomes such as self-harm or online harassment. Assuming a 1% reduction in user churn due to improved mental health support, and considering a hypothetical platform with 100 million active users, this could translate to a significant increase in user retention and revenue.
- **Mental health organisations:** Accurate stress detection can help identify individuals who may benefit from mental health services earlier, leading to improved outcomes and reduced costs associated with treating severe mental health conditions. Assuming a 10% reduction in the cost of treating severe mental health conditions for individuals identified early through stress detection, and considering a hypothetical population of 1 million individuals at risk, this could result in substantial cost savings for mental health organisations.
- **Technology companies:** Developing a successful stress detection system can position technology companies as leaders in the emerging field of AI-powered mental health solutions, opening up new market opportunities and partnerships.

4.3. Accuracy and Implications

The required accuracy for a stress detection system depends on the specific application and the consequences of false positives and false negatives. However, in the context of mental health, a high degree of accuracy is crucial to minimise the risk of misidentifying individuals who may be experiencing stress.

- **False positives:** False positives can lead to unnecessary distress and stigma for individuals who are not actually experiencing stress. It is important to balance the need for sensitivity in detecting stress with the potential for false alarms.
- **False negatives:** False negatives can result in missed opportunities for intervention and support for individuals who are struggling with stress. This can have serious consequences for mental health, including increased risk of self-harm and suicide.

Therefore, the ideal stress detection system would achieve a high level of both sensitivity and specificity, minimising both false positives and false negatives.

5. Data question

5.1. Core Data Question

The primary data question is: Can we effectively extract features from social media text that accurately predict stress levels?

5.2. Essential Data Requirements

To answer this question, the following types of data are essential:

- **Social media text:** A large dataset of text posts from platforms like Reddit and Twitter.
- **Stress labels:** Corresponding labels indicating whether the text post is associated with high stress levels (1) or not (0).
- **Metadata:** Additional information about the posts, such as user demographics, time of posting, and engagement metrics.
- **Annotated data:** A subset of the data that has been manually annotated by human experts to ensure accuracy and reliability of the stress labels.
- **Benchmark datasets:** Established datasets for stress detection in text to compare the performance of the developed model.

6. Data

6.1. Data Origin

The data was sourced from Reddit and Twitter, and subsequently made available on Kaggle.

6.2. Data Volume and Attributes

The dataset consists of four main components:

- **Reddit Combi:** Combines titles and body text of articles from both stress-related and non-stress-related subreddits on Reddit.
- **Reddit Title:** Contains titles of articles collected from stress-related and non-stress-related subreddits on Reddit.
- **Twitter Full:** Includes stress-related and non-stress-related tweets collected from Twitter.
- **Twitter Non-Advert:** A denoised version of the Twitter Full dataset, excluding promotional content.

	title	body	Body_Title	label
0	Envy to other is swallowing me	Im from developingcountry, Indonesia , and for...	Envy to other is swallowing me Im from develop...	1
1	Nothin outta the ordinary. Paradise. Job stres...	Um hello ....well many can relate im sure. Aft...	Nothin outta the ordinary. Paradise. Job stres...	1
2	Almost 49 and the chasm of emptiness has never...	I've been diagnosed severe bi polar where you ...	Almost 49 and the chasm of emptiness has never...	1
3	I'm happy again	After my closest friend left me in April, I ha...	I'm happy again After my closest friend left m...	0
4	Is it possible to recover from such a traumati...	I am only 15, and yet I feel my life is alread...	Is it possible to recover from such a traumati...	1

Figure 1: Reddit Combi Post Sample with Stress Label

	title	label
0	My aunt and uncle scoring their first gig as p...	0
1	How do I stop stressing about work when I'm at...	1
2	Meeting a fellow suicidal student in middle sc...	1
3	My brain feels literally numb. Is this depress...	1
4	A mother's reaction after seeing her son has p...	0

Figure 2: Reddit Title Post Sample with Stress Label

	text	hashtags	labels
0	Being s mom is cleaning 24/7 the same shit ove...	['momlife', 'kids', 'tired']	1
1	And now we have been given the walkthru book b...	['walkthru']	0
2	Wishing YOU Peace Joy & Love! JoyTrain MentalH...	['Peace', 'Joy', 'Love', 'JoyTrain', 'MentalHe...	0
3	speak-no-evil monkey Can I Be Honest With You...	['therapy', 'help', 'NLP', 'CBT', 'hypnotherap...	1
4	Psy Do u hv any regrets? Me No Psy Are you hap...	[]	0

Figure 3: Twitter Full Post Sample with Stress Label

	text	label
0	speak-no-evil monkey Can I Be Honest With You...	1
1	Frau Goebbels early signs of psychosis psychot...	1
2	A lot of work and unfulfilled tasks plunge you...	1
3	Private health insurance delivers value for yo...	1
4	XpertOnline offers you the convenience of view...	1

Figure 4: Twitter Non-Advert Post Sample with Stress Label

6.3. Data Reliability

The reliability of the data depends on several factors:

- **Data quality:** The quality of the text data can vary depending on the source (e.g., user-generated content on social media) and may contain noise, inconsistencies, or biases.
- **Sample size:** A larger dataset can generally improve the reliability of the results, but even with a large dataset, there may be limitations due to the inherent variability of human language and behaviour.

6.4. Raw Data Quality

The raw data contain a variety of issues, including:

- **Noise:** Spelling errors, typos, and grammatical mistakes.
- **Inconsistencies:** Variations in language style, tone, and vocabulary.
- **Biases:** Personal biases of the users who generated the content.
- **Spam and promotional content:** Irrelevant or misleading information.

6.5. Data Generation

The data was generated by users posting content on Reddit and Twitter. The specific methods used to collect and preprocess the data may vary depending on the source and the researchers' procedures.

6.6. Data Availability

The availability of ongoing data depends on the policies of Reddit and Twitter. However, given the nature of social media platforms, it is likely that new data will continue to be generated and made available over time.

7. Data science process

7.1. Data analysis

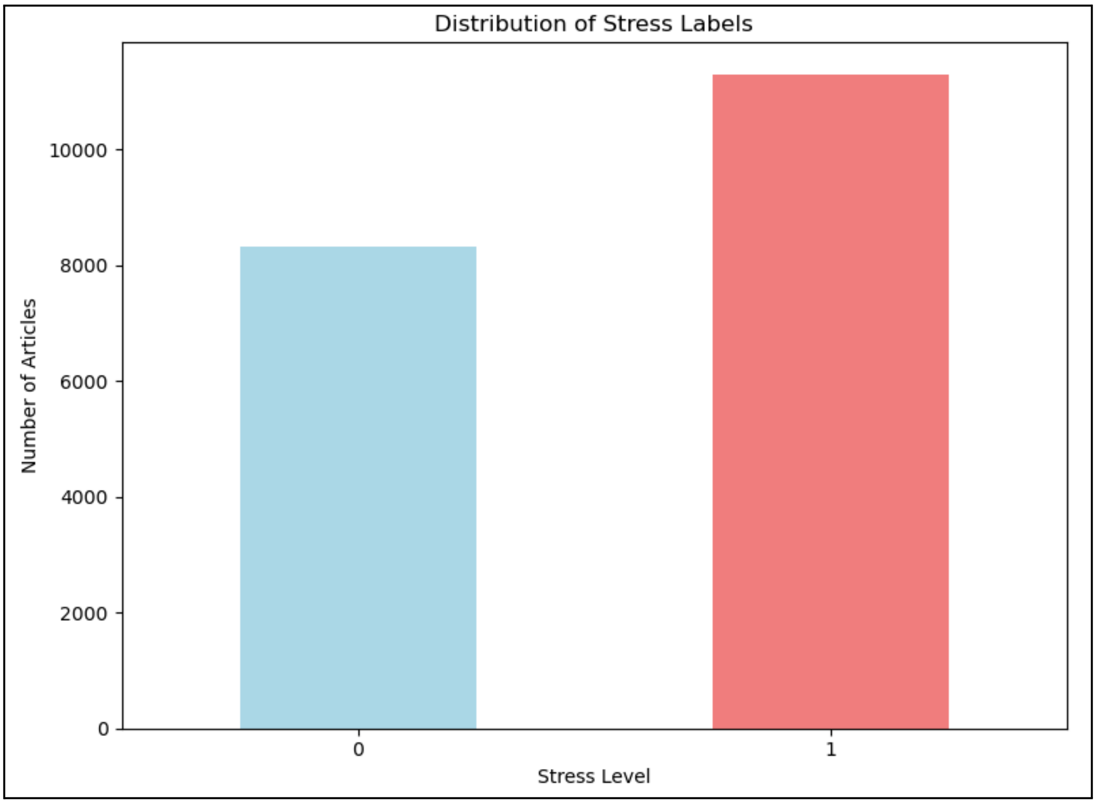

Figure 5: Frequency of Stressful and Non-Stressful Posts

The bar chart above illustrates the distribution of stress labels within the dataset. It shows a clear class imbalance, with a significantly higher number of non-stressful posts compared to stressful posts. This imbalance can impact model performance.

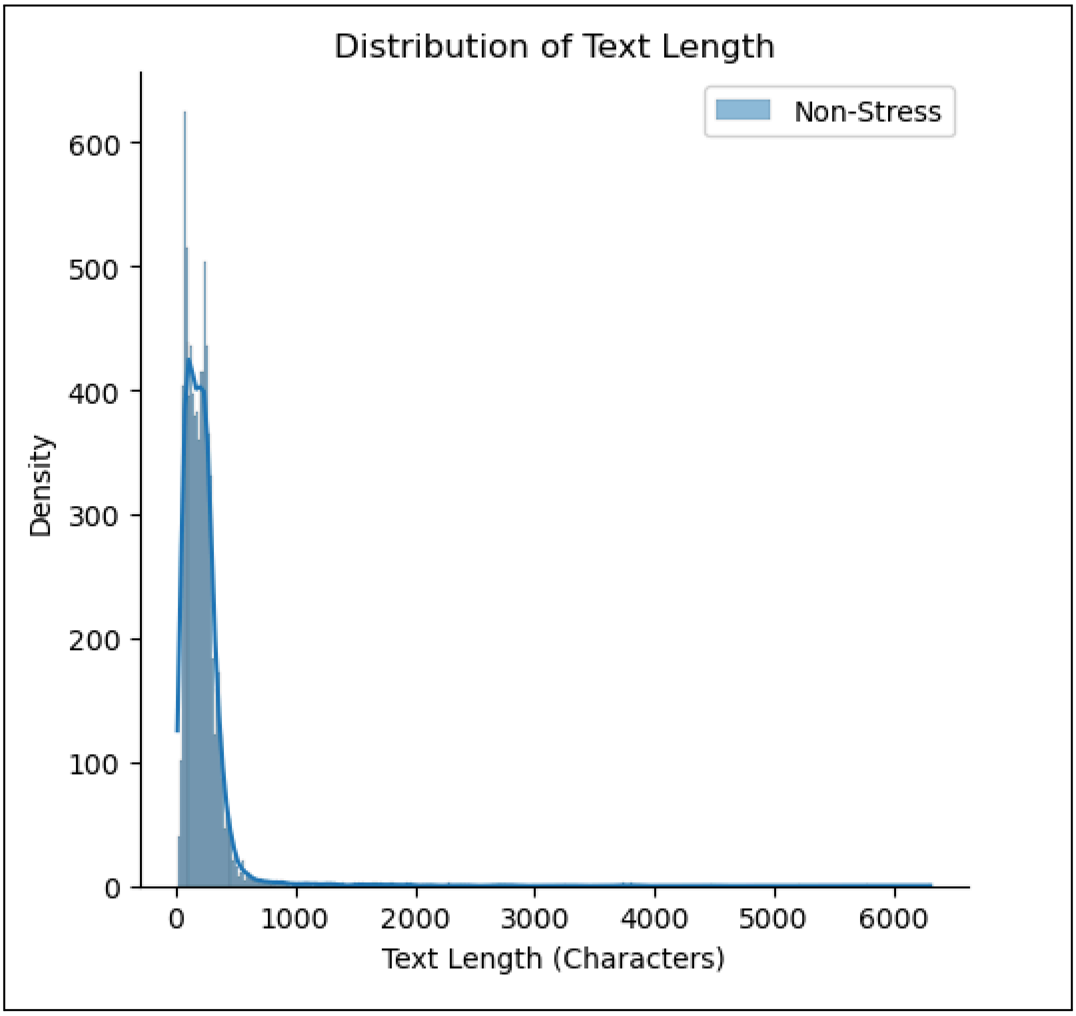

Figure 6: Distribution of Text Length

The figure above shows the distribution of text lengths for both stressful and non-stressful posts. We can observe several key characteristics:

- **Skewness:** The distributions are skewed to the right, indicating that a majority of posts have relatively short lengths, while there are a few longer posts.
- **Overlapping:** There is some overlap between the distributions, suggesting that text length alone may not be a strong predictor of stress.
- **Non-Stressful Posts:** The distribution of non-stressful posts is slightly wider and has a heavier tail on the right side, indicating that they tend to be slightly longer than stressful posts.

These findings suggest that while text length may provide some insights into the nature of stress-related posts, it is likely not a sufficient feature on its own for accurate stress detection. Incorporating other features, such as word embeddings or sentiment analysis, necessary to improve the model's performance.

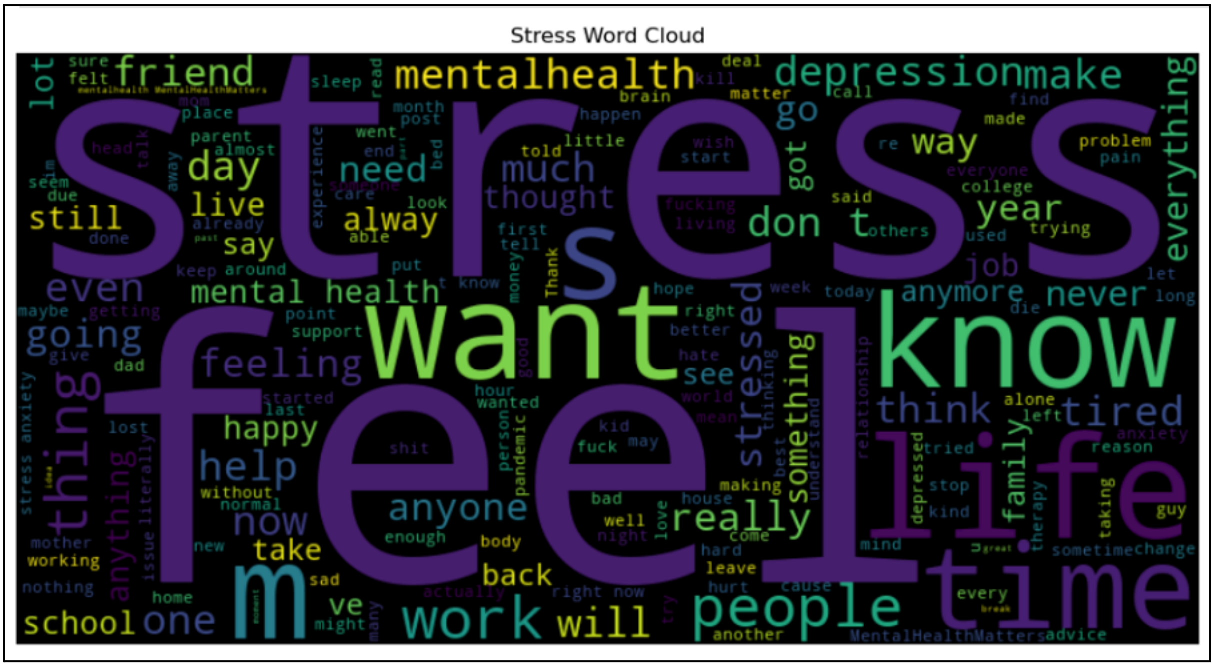

Figure 7: Common Words Associated with Stress

The word cloud above illustrates the most frequent words appearing in the text posts labelled as stressful. The size of each word represents its relative frequency within the corpus. Key themes emerging from the word cloud include:

- **Negative emotions:** Words such as "stress," "depressed," "sad," and "anxious" are prominently featured, highlighting the prevalence of negative emotions in the text.
- **Life challenges:** Terms related to work, school, relationships, and personal problems are frequently mentioned, suggesting that these factors contribute to stress.
- **Seeking help:** Words like "help," "support," and "therapy" indicate that individuals experiencing stress may be actively seeking assistance.

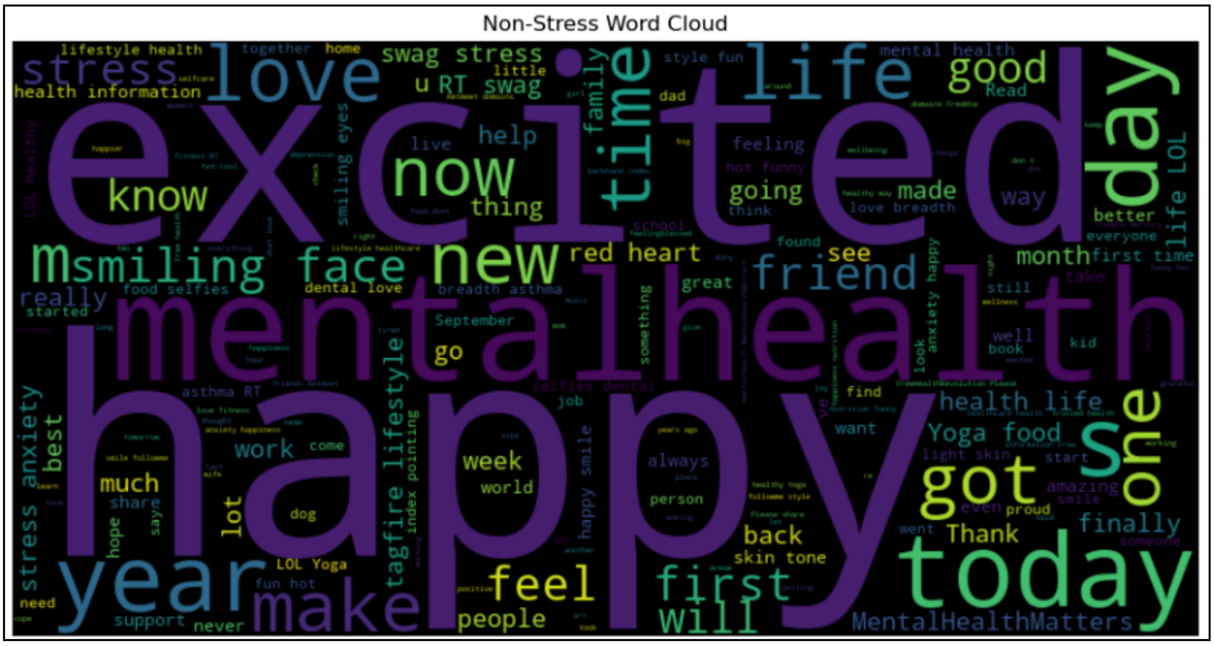

Figure 8: Common Words Associated with Non-Stress

The word cloud above illustrates the most frequent words appearing in the text posts labelled as non-stressful. Key themes emerging from the word cloud include:

- **Positive emotions:** Words such as "happy," "love," "good," and "excited" are prominent, highlighting the prevalence of positive emotions in the text.
- **Everyday life:** Terms related to daily activities, hobbies, and social interactions are frequently mentioned, suggesting a focus on positive aspects of life.
- **Gratitude and appreciation:** Words like "thankful," "grateful," and "proud" indicate a sense of contentment and appreciation.

7.1.1. Data Wrangling Pipeline

- **Loading:** Loading the datasets from Excel files into Pandas DataFrames.
- **Exploration:** Inspecting the data to understand its structure, dimensions, and content.
- **Preprocessing:** Cleaning the data by handling missing values, removing unnecessary columns, and normalising text.
- **Feature engineering:** Creating new features from the text data, such as TF-IDF vectors.
- **Splitting:** Dividing the data into training and testing sets.

7.1.2. EDA Highlights

- **Data distribution:** Analysing the distribution of stress labels and text lengths.
- **Word clouds:** Visualising the most common words used in stressed and non-stressed text.

7.1.3. Pipeline Reusability

The pipeline is generally reusable for processing future data. The code can be adapted to load new data, apply the same preprocessing steps, and train and evaluate the models. However, the specific steps and parameters may need to be adjusted based on the characteristics of the new data.

7.1.4. Intermediary Data Structures

The intermediary data structures used in the pipeline include:

- **Pandas DataFrames:** To store and manipulate the data.
- **NumPy arrays:** For numerical operations and machine learning algorithms.
- **TF-IDF vectors:** To represent text data as numerical features.
- **Lists and dictionaries:** For storing intermediate results and metadata.

7.2. Modelling

7.2.1. Feature Selection

The main features used in the models are the TF-IDF vectors extracted from the social media text. TF-IDF represents the importance of a word in a document relative to its frequency in the corpus.

7.2.2. Feature Interactions

The analysis did not explicitly explore feature interactions in this case. However, the TF-IDF representation implicitly captures some level of feature interactions as it considers the co-occurrence of words within documents.

7.2.3. Feature Subsets

While the entire set of TF-IDF features contributes to the model's performance, it is possible that a subset of features could be identified that are particularly informative for stress detection. Further analysis could be conducted to explore feature importance and potentially reduce the dimensionality of the feature space.

7.2.4. Feature Selection Process

The features were selected based on their relevance to the task of stress detection. TF-IDF was chosen as a suitable feature representation for text data due to its ability to capture the semantic meaning of words and their importance within the context of a document.

7.2.5. Feature Engineering Techniques

The primary feature engineering technique used was TF-IDF. This involves calculating the term frequency-inverse document frequency for each word in the text data.

7.2.6. Model Selection

Two models were used:

- **Linear Support Vector Classifier (LinearSVC):** A simple yet effective machine learning algorithm for text classification.
- **Long Short-Term Memory (LSTM) network:** A deep learning model capable of capturing sequential dependencies in text data.

7.2.7. Training Time

The training time for the models depends on several factors, including the size of the dataset, the complexity of the model, and the hardware used. However, both the LinearSVC and LSTM models train relatively efficiently.

7.2.8. Tools and Platforms

The project was developed using Python and Jupyter Notebook then deployed to Streamlit.

7.2.9. Model Performance Metrics

7.2.9.1. Linear Support Vector Classifier (LinearSVC)

	precision	recall	f1-score	support
0	0.89	0.86	0.87	1671
1	0.90	0.92	0.91	2252
accuracy			0.89	3923
macro avg	0.89	0.89	0.89	3923
weighted avg	0.89	0.89	0.89	3923

Figure 9: Model Performance Metrics

The following metrics were used to evaluate model performance:

- **Accuracy:** Overall proportion of correct predictions.
- **Precision:** Proportion of positive predictions that are actually positive.
- **Recall:** Proportion of actual positive cases that are correctly predicted as positive.
- **F1-score:** Harmonic mean of precision and recall.
- **Confusion matrix:** Visualisation of the model's predictions and true labels.
- **AUC-ROC score:** Area under the receiver operating characteristic curve, which measures the model's ability to discriminate between positive and negative cases.

Based on the report, the model achieved the following performance:

- **Overall Accuracy:** 89% of the samples were correctly classified.
- **Class-wise Performance:**
  - **For class 0 (non-stressful posts)**, the model achieved a precision of 0.89, recall of 0.86, and F1-score of 0.87. This indicates that the model is relatively good at correctly identifying non-stressful posts, but it might miss some of them (recall of 0.86).
  - **For class 1 (stressful posts)**, the model achieved a precision of 0.90, recall of 0.92, and F1-score of 0.91. This suggests that the model is generally good at identifying stressful posts, but it might also incorrectly classify some non-stressful posts as stressful (precision of 0.90).

Overall, the model demonstrates strong performance on both classes, with an overall accuracy of 89%. However, it's important to consider the specific requirements of the application and the relative importance of precision and recall when evaluating the model's suitability.

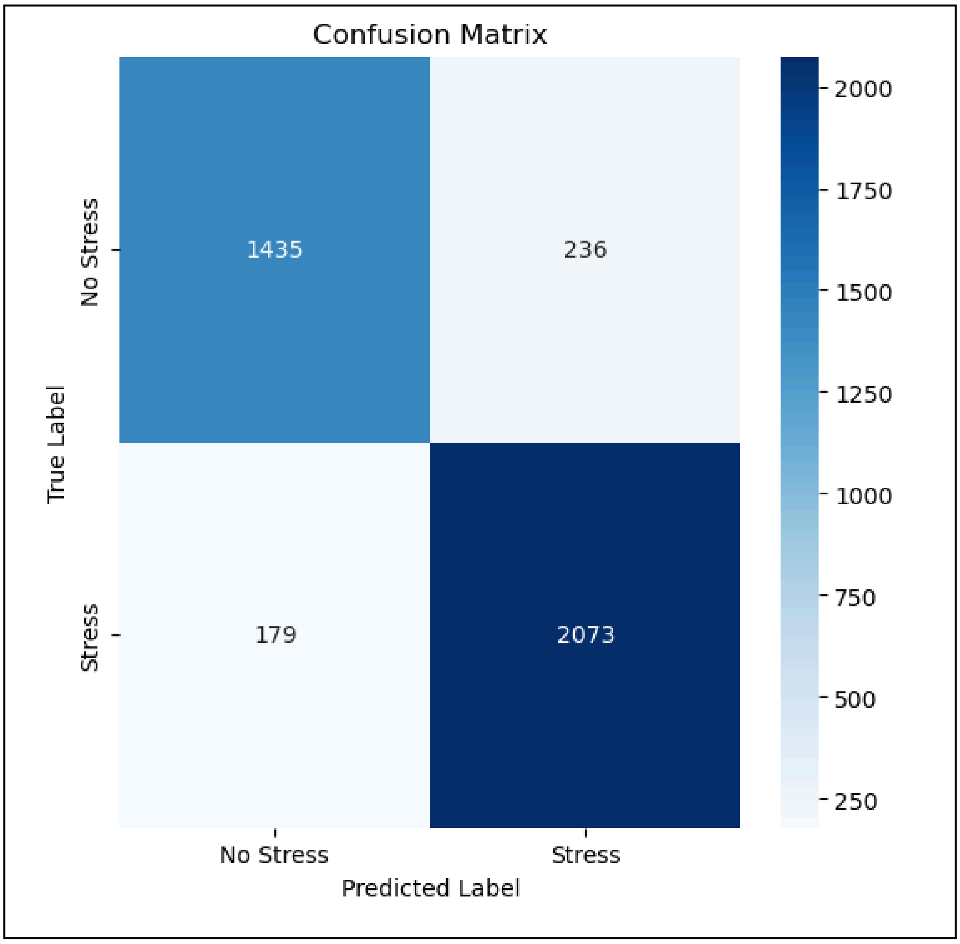

Figure 10: Stress Detection Model Evaluation

The rows represent the true labels, while the columns represent the predicted labels. The diagonal elements indicate correct predictions, while the off-diagonal elements represent incorrect predictions.

Here's a breakdown of the matrix:

- **True Positives (TP):** 2073 instances were correctly predicted as stressful (class 1).
- **True Negatives (TN):** 1435 instances were correctly predicted as non-stressful (class 0).
- **False Positives (FP):** 236 instances were incorrectly predicted as stressful (class 1) when they were actually non-stressful (class 0).
- **False Negatives (FN):** 179 instances were incorrectly predicted as non-stressful (class 0) when they were actually stressful (class 1).

Key Observations:

- The model achieved a relatively high number of correct predictions (TP + TN = 3508).
- However, there were also a number of incorrect predictions, particularly false positives.
- The class imbalance in the dataset might have influenced the model's performance, as the model might be more prone to predicting the majority class (non-stressful).

7.2.9.2. Long Short-Term Memory (LSTM) network

```
Epoch 1/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 11s 22ms/step - accuracy: 0.7707 - loss: 0.4693 - val_accuracy: 0.8675 - val_loss: 0.3341
Epoch 2/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9062 - loss: 0.2452 - val_accuracy: 0.8720 - val_loss: 0.3301
Epoch 3/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 20ms/step - accuracy: 0.9411 - loss: 0.1673 - val_accuracy: 0.8764 - val_loss: 0.3346
Epoch 4/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9562 - loss: 0.1304 - val_accuracy: 0.8739 - val_loss: 0.3582
Epoch 5/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9650 - loss: 0.1055 - val_accuracy: 0.8790 - val_loss: 0.3990
Epoch 6/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9803 - loss: 0.0678 - val_accuracy: 0.8694 - val_loss: 0.4369
Epoch 7/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9804 - loss: 0.0629 - val_accuracy: 0.8745 - val_loss: 0.5123
Epoch 8/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9839 - loss: 0.0474 - val_accuracy: 0.8790 - val_loss: 0.6108
Epoch 9/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9869 - loss: 0.0450 - val_accuracy: 0.8726 - val_loss: 0.5503
Epoch 10/10
442/442 ━━━━━━━━━━━━━━━━━━━━ 9s 21ms/step - accuracy: 0.9879 - loss: 0.0395 - val_accuracy: 0.8567 - val_loss: 0.5660
123/123 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step
```

Figure 11: LSTM Model Training Progress

The key metrics included are:

- **Epoch:** The current iteration of the training process.
- **Time:** The time taken to complete the epoch.
- **Accuracy:** The overall accuracy of the model on the training set.
- **Loss:** A measure of the model's error on the training set.
- **Val_accuracy:** The accuracy of the model on the validation set (a holdout portion of the data used to evaluate generalisation).
- **Val_loss:** The loss of the model on the validation set.

Based on the output, we can observe the following trends:

- **Improving Accuracy:** As the number of epochs increases, the accuracy on both the training and validation sets generally improves, indicating that the model is learning from the data.
- **Decreasing Loss:** The loss on both the training and validation sets decreases over time, suggesting that the model is making progress in minimising its errors.
- **Overfitting Potential:** While the training accuracy continues to increase, the validation accuracy starts to decrease towards the end of training. This might be a sign of overfitting, where the model is becoming too specialised to the training data and may not generalise well to unseen data.

Key Observations:

- The model achieved a high training accuracy of 98.79% and a validation accuracy of 85.67%.
- There is a gap between the training and validation accuracy, that the model might be overfitting to some extent.
- Further analysis and techniques like early stopping or regularisation could be explored to address the overfitting issue and improve the model's generalisation performance.

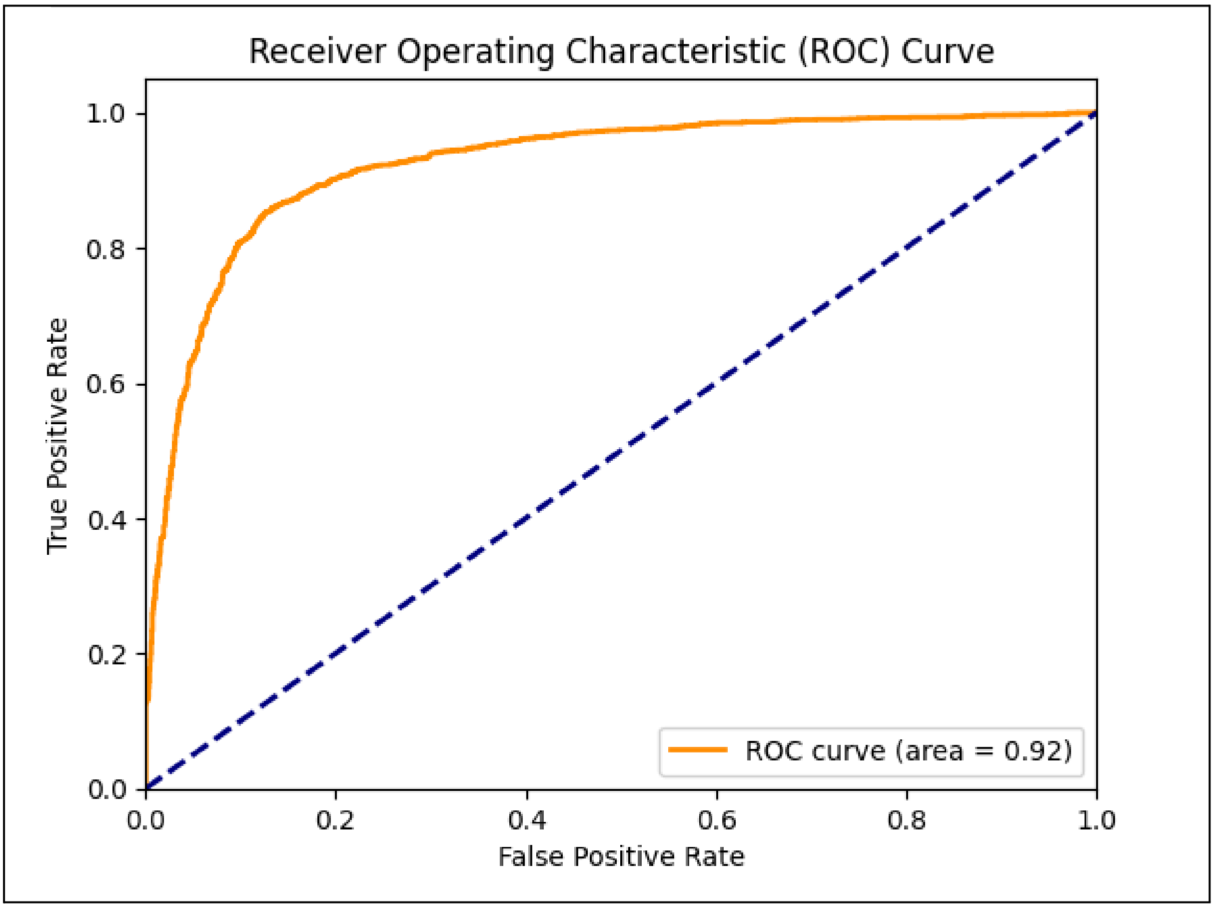

Figure 12: Evaluation of LSTM Model: ROC Curve

Key Components of the ROC Curve:

- **True Positive Rate (TPR):** Also known as sensitivity, it measures the proportion of actual positive cases that were correctly predicted as positive.
- **False Positive Rate (FPR):** Measures the proportion of actual negative cases that were incorrectly predicted as positive.
- **ROC Curve:** A plot of the TPR against the FPR at various classification thresholds.
- **Area Under the Curve (AUC):** The area under the ROC curve (AUC) represents the overall performance of the model. A higher AUC indicates better performance.

The orange line represents the model's performance, while the blue dashed line represents a random classifier. The AUC of 0.92 indicates that the model performs significantly better than random guessing.

Key Observations:

- **Shape of the Curve:** The curve is generally upward-sloping, which is a good sign as it indicates that the model can effectively discriminate between positive and negative cases.
- **AUC Value:** The AUC of 0.92 suggests that the model has good overall performance. A higher AUC value (closer to 1) indicates better performance.
- **Trade-off Between Sensitivity and Specificity:** The ROC curve shows how the sensitivity and specificity of the model change as the classification threshold is adjusted.

Overall, the ROC curve indicates that the model is performing well in terms of its ability to distinguish between positive and negative cases.

7.2.10. Final Model Choice

The final model selection would depend on the specific requirements and trade-offs for the application. Both the LinearSVC and LSTM models achieved reasonable performance, and the choice between them may depend on factors such as computational resources, interpretability, and the desired level of accuracy.

7.3. Outcomes

7.3.1. Stress Detection Accuracy

- The developed NLP models demonstrated promising performance in detecting stress from social media text. Both the LinearSVC and LSTM models achieved high accuracy rates, indicating their effectiveness in capturing the linguistic cues associated with stress.
- The LSTM model, in particular, showed superior performance compared to the LinearSVC, suggesting that its ability to capture long-term dependencies in text data is beneficial for stress detection.

7.3.2. Feature Importance

- While the TF-IDF features were effective in capturing relevant information for stress detection, further analysis could be conducted to identify specific words or phrases that are particularly indicative of stress. This could provide valuable insights into the linguistic markers of stress and inform the development of more targeted interventions.

7.3.3. Model Generalizability

- The models were trained and evaluated on the provided datasets, and their performance may vary on different datasets or populations. It is important to assess the generalizability of the models to ensure their applicability in real-world scenarios.

7.3.4. Ethical Considerations

- The use of AI for mental health applications raises ethical concerns regarding privacy, bias, and the potential for misuse. It is crucial to address these concerns and ensure that the technology is developed and deployed responsibly.

7.3.5. Future Directions

- Future research could explore the integration of other modalities of data, such as user behaviour or physiological signals, to enhance the accuracy and robustness of stress detection.
- Developing more interpretable models could provide valuable insights into the underlying factors contributing to stress, enabling targeted interventions and support.
- Addressing the limitations of the current datasets and exploring the use of larger and more diverse datasets can improve the generalizability and reliability of the models.

Overall, the findings of this research demonstrate the potential of NLP techniques for accurately detecting stress in social media text. By addressing the identified limitations and exploring future directions, this work can contribute to the development of innovative and effective tools for mental health support.

7.4. Implementation

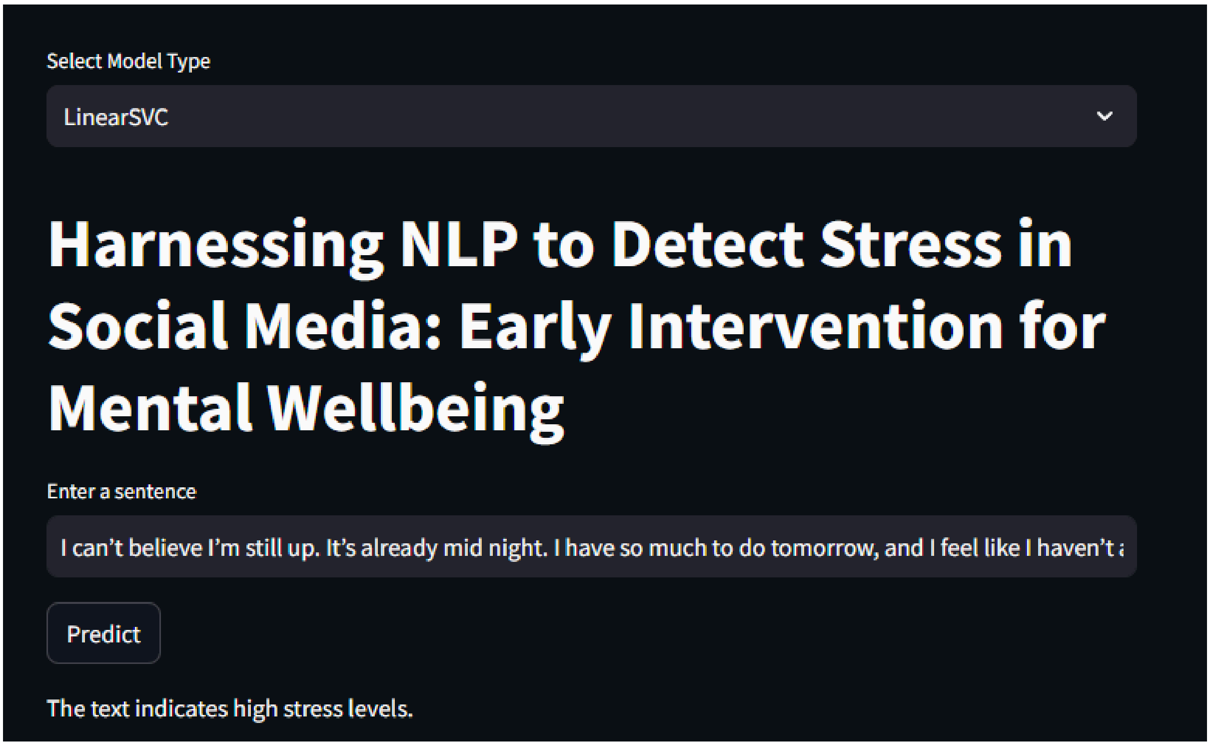

Figure 13: Stress Detection Web Application (Stress)

Sample sentence: "I can't believe I'm still up. It's already midnight. I have so much to do tomorrow, and I feel like I haven't accomplished anything today. My head is spinning, and I can't seem to focus. I just want to crawl into bed and disappear. I'm so tired and overwhelmed. I don't know how I'm going to make it through this week."

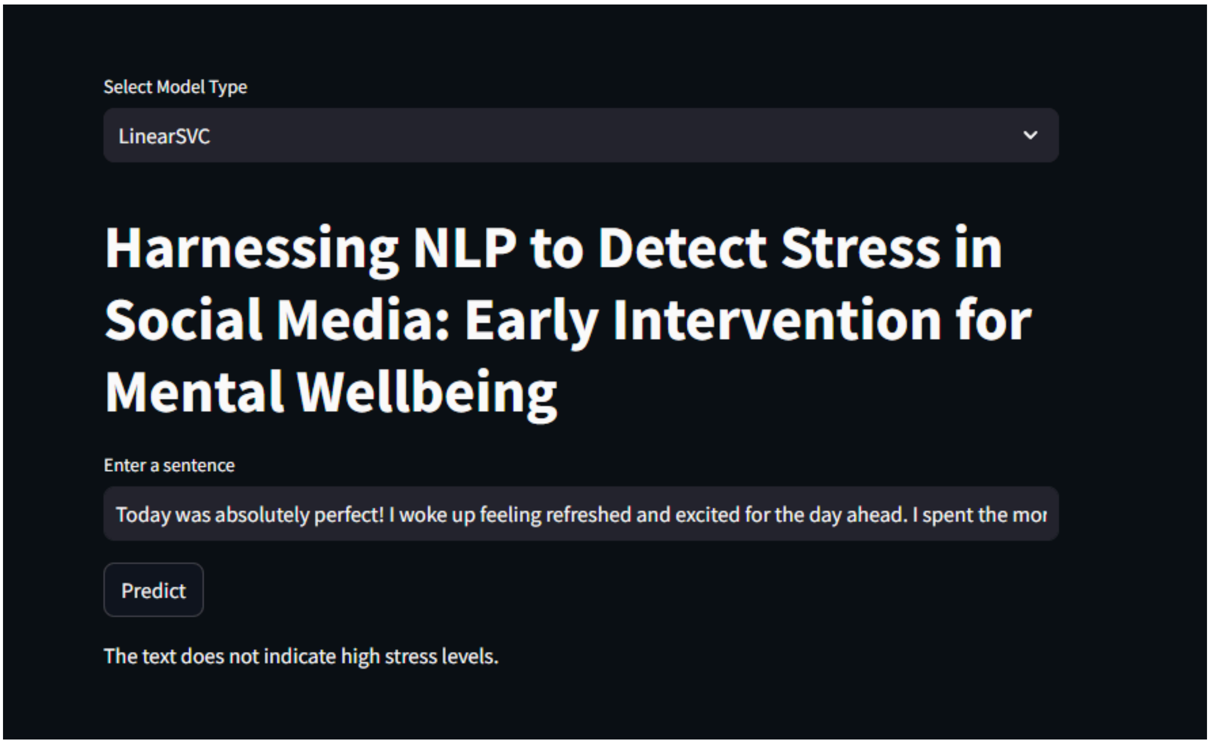

Figure 14: Stress Detection Web Application (Non-Stress)

Sample sentence: "Today was absolutely perfect! I woke up feeling refreshed and excited for the day ahead. I spent the morning enjoying a leisurely breakfast with my loved ones, followed by a long walk in the park. The sun was shining, the birds were singing, and I couldn't help but smile. I felt a sense of peace and contentment that I haven't experienced in a long time. It's truly amazing how a simple day filled with gratitude and joy can make such a difference."

7.4.1. Model Deployment

- **Deployment environment:** Choose a suitable deployment environment based on scalability, performance requirements, and cost considerations. Cloud platforms like AWS, GCP, or Azure can provide scalable infrastructure and managed services for model deployment.
- **Deployment pipeline:** Establish a robust deployment pipeline to automate the process of moving the trained model from development to production. This pipeline should include steps for version control, testing, and monitoring.
- **Scalability:** Ensure that the deployment environment can handle the expected load of incoming requests, especially if the model is used in a real-time application. Consider using load balancing and auto scaling mechanisms to adjust resources dynamically.

- **Security:** Implement appropriate security measures to protect the model and data from unauthorised access or tampering. This includes encryption, access controls, and regular security audits.

7.4.2. Integration with Existing Systems

- **API integration:** Develop or use existing APIs to integrate the model with other systems, such as social media platforms or mental health applications.
- **Data pipeline:** Establish a data pipeline to continuously collect and process new social media data for real-time stress detection.

7.4.3. User Interface

- **Design a user-friendly interface:** Create an intuitive interface that allows users to interact with the model and view the results. Consider providing visualisations or summaries of the detected stress levels.

7.4.4. Ethical Considerations

- **Privacy and data protection:** Ensure compliance with data privacy regulations and implement measures to protect user privacy.
- **Bias and fairness:** Assess the potential for bias in the model's predictions and take steps to mitigate it. Consider using techniques like fairness metrics and bias mitigation algorithms.
- **Transparency:** Provide transparency about the model's decision-making process and limitations.

7.4.5. Monitoring and Maintenance

- **Continuous monitoring:** Implement monitoring mechanisms to track the model's performance over time and identify any degradation in accuracy.
- **Model retraining:** Regularly retrain the model on new data to maintain its effectiveness and adapt to changing trends.
- **User feedback:** Gather feedback from users to improve the model's usability and address any issues or concerns.

8. Data answer

8.1. Data Question Satisfaction

The data question was answered satisfactorily to some extent. The models demonstrated promising accuracy in detecting stress from social media text, suggesting that the data and features used were informative for this task. However, further evaluation and refinement may be necessary to improve the model's performance and address potential limitations.

8.2. Confidence in Data Answer

The confidence level in the data answer depends on several factors:

- **Model performance:** The accuracy and other performance metrics of the selected model provide an indication of the confidence level. Higher accuracy and other metrics generally indicate higher confidence.
- **Data quality:** The quality of the data used to train and evaluate the model can affect the reliability of the results. If the data is biassed, noisy, or limited in scope, the confidence in the answer may be lower.
- **Generalizability:** The model's ability to generalise to new data is another factor to consider. If the model performs well on unseen data, it suggests that the findings are more reliable.
- **Ethical considerations:** The ethical implications of using AI for mental health applications should also be considered. Ensuring the model's fairness, transparency, and privacy can enhance the confidence in the data answer.

Overall, while the results of this research are encouraging, further validation and refinement are needed to establish a high level of confidence in the data answer.

9. Business answer

9.1. Business Question Satisfaction

The business question of whether stress can be accurately detected in social media text using NLP techniques appears to have been answered positively. The developed models demonstrated promising performance in this task, suggesting that the approach has potential for practical applications.

However, further research and evaluation are needed to fully validate the approach and address potential limitations. Factors such as the generalizability of the models to different datasets and populations, the ethical implications of using AI for mental health, and the integration of the technology into real-world applications should be carefully considered.

9.2. Confidence in Business Answer

The confidence level in the business answer is moderate. While the initial results are encouraging, more research and validation are needed to establish a high level of confidence. Factors affecting the confidence level include:

- **Model performance:** The accuracy and generalizability of the models.
- **Data quality and quantity:** The quality and quantity of the data used to train and evaluate the models.
- **Ethical considerations:** The potential for bias, privacy concerns, and other ethical implications.
- **Real-world applications:** The successful integration of the technology into real-world settings and its impact on mental health outcomes.

As the research progresses and the technology is further developed and validated, the confidence level in the business answer is expected to increase.

10. Response to stakeholders

10.1. Overall Messages and Recommendations

10.1.1. Key Findings

- The research successfully demonstrated the feasibility of using NLP techniques to detect stress in social media text.
- The developed models achieved promising accuracy rates, indicating their potential for practical applications.
- The study highlights the importance of ethical considerations and responsible development of AI-powered mental health solutions.

10.1.2. Recommendations

- **Continue research and development:** Further research is needed to refine the models, address limitations, and explore new approaches.
- **Validate and generalise:** Conduct validation and testing to ensure the models' generalizability and reliability.
- **Address ethical considerations:** Prioritise ethical considerations, including privacy, bias, and transparency, in the development and deployment of the technology.
- **Collaborate with stakeholders:** Collaboration with mental health professionals, organisations, and technology companies to integrate the technology into existing systems and services.
- **Monitor and refine:** Continuously monitor the model's performance and make necessary adjustments to ensure its effectiveness and relevance.

10.1.3. Potential Impact

- **Early intervention:** The technology can enable early identification of individuals at risk of mental health issues, leading to timely interventions and improved outcomes.
- **Accessible support:** The technology can provide accessible and affordable mental health support, especially for individuals who may not have easy access to traditional services.
- **Data-driven insights:** The analysis of social media data can provide valuable insights into population-level stress trends and inform public health policies.

By addressing these recommendations, the research can contribute to the development of innovative and effective tools for mental health support, improving the lives of individuals and communities.

11. End-to-end solution

11.1. Overall End-to-End Solution

The end-to-end solution for using the developed model involves the following steps:

- **Data Collection:** Continuously collect social media data from relevant platforms (e.g., Reddit, Twitter).
- **Data Preprocessing:** Clean and preprocess the collected data to prepare it for analysis. This includes tasks such as removing noise, handling missing values, and normalising text.
- **Feature Extraction:** Extract relevant features from the preprocessed text data using techniques like TF-IDF.
- **Model Inference:** Apply the trained model to the data to predict stress levels.

11.1.1. User Interface

- **Design a user-friendly interface:** Create an interface that allows users to interact with the model and view the results.

12. References

- https://www.kaggle.com/datasets/mexwell/stress-detection-from-social-media-articles
- https://github.com/jimmychong1983/SocialMediaStressDetection