Anxiety Detection in Reddit Posts Through Emotion Dynamics Analysis

Ashala Senanayake[a] and Zilu Liang[b]
Ubiquitous and Personal Computing Lab, Kyoto University of Advanced Science (KUAS), Japan

Keywords: Anxiety Detection, Emotional Dynamics, Social Media, Natural Language Processing.

Abstract: Anxiety disorders have become a significant subset of mental health challenges in the context of complex modern social life. The widespread integration of social media into daily life has created platforms for individuals to share their updates, offering a rich resource for linguistic and behavioural analysis outside traditional clinical settings. Among these platforms, Reddit stands out as a valuable tool for researchers due to its rich and diverse textual data. This paper leverages five commonly used machine learning models: decision tree, random forest (RF), k-nearest neighbours, linear regression, and naive Bayes to explore the emotional dynamics present in Reddit posts for detecting anxiety. Reddit posts from 1,800 users, categorized as either anxiety or non-anxiety, were used for model training, validation, and testing, with data split into 70%, 15%, and 15%, respectively. The decision-making process of the best-performing model was evaluated by incorporating feature importance. RF achieved the best performance among all models, with an accuracy of 89%. Its interpretation revealed that average emotion scores and normalized emotion gaps were key factors, highlighting the significance of emotional intensity and variability over time. Furthermore, the results indicate that emotions such as sadness and joy play a particularly significant role in detecting anxiety.

1 INTRODUCTION

The increase in socio-economic adversities has led to a rise in mental health issues among the public. Anxiety disorder occupies a major subset among these mental issues. Generally, anxiety disorders are identified as persistent feeling of worry, dread when gets out of control and interfere with daily life (Anxiety disorders, 2024). According to the National Comorbidity Survey Adolescent Supplement (NCS-A), an estimated lifetime prevalence of any anxiety disorder among U.S. adolescents aged 13-18 is 31.9% and estimated 8.3% had severe impairment. Furthermore, prevalence of any anxiety disorder among adolescents was higher for females (38.0%) than for males (26.1%) (Kessler, Chiu, Demler, & Walters, 2005).

Social media penetration has resulted in the exponential growth of social media engagement, allowing people to share personal updates such as achievements, celebrations, vacations, and, most importantly, feelings. These posts can take the form of images, sounds, texts, or a combination of them. Analysing social media texts provides valuable insights into someone's mental state (Burke, Marlow, & Lento, 2010). Researchers' attention has drawn to identification various insights from social media texts. Reddit is such a popular social media platform among researchers who are interested in linguistic analysis. It has become an important resource for researchers because of its varied and engaged user base, extensive content, and the well-organized structure of its communities (De Choudhury & De, 2014). The platform offers researchers a unique perspective on social interactions making it a popular tool for research in areas such as sociology, psychology, marketing, and data science.

Prior studies have analysed Reddit posts to identify linguistic markers related to mental health, such as uni-grams, bi-grams, and tri-grams among college students (Bagroy, Kumaraguru, & De Choudhury, 2017). This study primarily focuses on linguistic and temporal patterns associated with academic, social, and personal stressors. However, it

[a] https://orcid.org/0009-0005-4634-1657
[b] https://orcid.org/0000-0002-2328-5016

Senanayake, A. and Liang, Z.
Anxiety Detection in Reddit Posts Through Emotion Dynamics Analysis.
DOI: 10.5220/0013375900003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 920-927
ISBN: 978-989-758-731-3; ISSN: 2184-4305

struggles to pinpoint specific emotional disturbances. Furthermore, Reddit data has been utilized to identify topics discussed during various mental health issues and how these topics evolve over time (Saha, Yousuf, Boyd, Pennebaker, & De Choudhury, 2022).

Previous studies have also developed explainable AI models using Reddit data for mental health disorder classification, but they lack longitudinal analysis, such as tracking users' language behaviour over time. As a result, these models fail to capture the dynamics of emotions in classification. This highlights a significant need to analyse emotional dynamics, particularly when focusing on specific emotional disturbances like anxiety (Kerz, Zanwar, Qiao, & Wiechmann, 2023).

To address the research gap, this study aims to analyse the emotional dynamics of anxious users and uncover the influence of emotions such as sadness, fear, anger, and joy on anxiety. Based on this, our research seeks to answer the following questions: How can anxiety be classified based on basic emotional dynamics? and What is the impact of different emotional dynamics on anxious users?

We adopted shallow machine learning techniques, including decision trees, random forests, and KNN, because they perform well with smaller datasets and are more interpretable compared to deep learning models, which is crucial in fields like medicine. Moreover, these models require less computational power and are often more robust to noisy data compared to deep learning models.

In the rest of the paper Section II describes the related works on the subjective topic, Section III presents the methodology incorporated including dataset used and models adopted. The results are covered in Section IV and finally Section V concludes the paper.

2 RELATED WORK

Anxiety has been detected through various approaches, including clinical, physiological, behavioural, emotional, and linguistic methods. Clinical diagnosis is typically performed by healthcare professionals through interviews and questionnaires. Physiological methods assess changes in heart rate, blood pressure, and hormone levels. Behavioural signs, such as restlessness and avoidance, offer further insight. Emotional indicators, such as mood swings and persistent worry, contribute to the identification of anxiety. Meanwhile, linguistic analysis focuses on speech patterns and word choices as potential signs of anxiety.

Linguistic detection of anxiety has been explored through several techniques. Sentiment analysis, for example, identifies the emotional tone in text (Saifullah, Fauziah, & Aribowo, 2021) (Senanayake, Priyadarshana, & Ranathunga, 2018), while emotion analysis delves into specific feelings (Hasan, Rundensteiner, & Agu, 2019). Language and vocabulary analysis examines word choice, repetition, and pronoun usage to uncover anxiety markers (Hofmann, Moore, Gutner, & Weeks, 2012). Additionally, topic modelling identifies recurring themes associated with anxiety (Shen & Rudzicz, 2017). Traditional machine learning models, such as regression and Naïve Bayes classification (Bagroy, Kumaraguru, & De Choudhury, 2017) (Lee, Sohn, & Choi, 2019), have been foundational in this area, while more recent deep learning-based methods offer a deeper understanding of natural language.

The literature highlights several emotion-annotated corpora as key resources for linguistic emotion analysis. Examples include the EmoBank (Buechel & Hahn, 2022), ISEAR (Scherer & Wallbott, 1994), and the SemEval datasets (2007, 2018, 2019) (Strapparava & Mihalcea, 2007) (Mohammad, Bravo-Marquez, Salameh, & Kiritchenko, 2018) (Chatterjee, Narahari, Joshi, & Agrawal, 2019). Lexicons like SenticNet (Biagioni, 2016), SentiWordNet (Baccianella, Esuli, Sebastiani, & others, 2010), and EmoSenticNet (Shah, Reyadh, Shaafi, Ahmed, & Sithil, 2019) are also instrumental in identifying emotions through keyword or lexical approaches.

Social media platforms, such as Twitter, Facebook, Sina Weibo, and VKontakte, have been valuable for psycholinguistic analysis (Lee, Sohn, & Choi, 2019) (Qiao, Yan, & Wang, 2024) (Nosov, Kuznetsova, Stankevich, Smirnov, & Grigoriev, 2023). Reddit, in particular, has widely been used for identifying human anxiety (Bagroy, Kumaraguru, & De Choudhury, 2017) (Shen & Rudzicz, 2017). A notable example is the Self-Reported Mental Health Diagnoses (SMHD) dataset from Reddit, which is specifically designed for studying anxiety and other mental health disorders (Cohan, et al., 2018). But this dataset is annotated post level, and it does not contain emotion level annotation too.

Since anxiety is a chronic disorder, rather than focusing solely on metrics from a single post, it would be helpful to consider multiple posts from the user's past. This approach allows us to incorporate metrics based on emotional dynamics over time.

In our study we mainly focus on emotional dynamics associated linguistic analysis for anxiety detection using social media posts.

3 METHODOLOGY

This section discusses the methodology for detecting anxiety by analysing emotional dynamics within Reddit posts. The first subsection of the methodology explains the adopted dataset, the second covers the process of feature engineering and data normalization, and the final subsection describes the models applied for the analysis. Fig. 1 illustrates the procedure we apply for anxiety detection using emotion dynamics.

3.1 Dataset

The analysis was conducted by using subset of the Reddit dataset provided by (Guo, Sun, & Vosoughi, 2021). The dataset spans the time window of 2011–2019. This is a self-reported emotional dataset which has validated manually, and to ensure the inclusion of posts made prior to diagnosis, self-reports were used as a proxy. Furthermore, to minimize noise caused by comorbidity, users who reported being diagnosed with more than one emotional disorder were excluded from the study. For our study, we select posts from 1,800 anxiety-diagnosed users and 1,800 users without any reported diagnoses. Furthermore, the dataset consists of 326,242 anxiety posts and 324,631 non-anxiety posts. Each post is annotated with scores for joy, sadness, anger, and fear, incorporating the SemEval - Affect in Tweets corpus (Mohammad, Bravo-Marquez, Salameh, & Kiritchenko, 2018) and Ekman's basic emotion framework, which identifies emotions closely related to emotional disorders (Ekman, 1999). This dataset is utilized for training, validation, and testing of the machine learning models, with proportions of 70%, 15%, and 15%, respectively.

3.2 Feature Engineering and Data Normalization

As the foundation for feature engineering, the raw emotion scores of joy, sadness, anger, and fear for each post were considered. While these raw scores capture post-specific emotion annotations, they may overlook broader trends and variability over time. To address this, emotion aggregates are incorporated to provide a comprehensive snapshot of emotional tendencies and extremes characterizing a user's emotional state.

At the user level, we calculate the average emotion scores shown in (1) including average joy, average sadness, average anger, and average fear, to represent the overall emotional tone. N is the total number of posts for a given user.

$$Average\ Emotion = \frac{1}{N}\sum_{i=1}^{N} E_{post,\,emotion_i} \tag{1}$$

Additionally, to capture the lowest emotional points, the minimum emotion scores within the time window are included as follows: minimum joy score, minimum sadness score, minimum anger score, and minimum fear score. Similarly, to capture the peak emotional intensities, the maximum emotion scores were calculated within the same time window as: maximum joy score, maximum sadness score, maximum anger score, and maximum fear score.

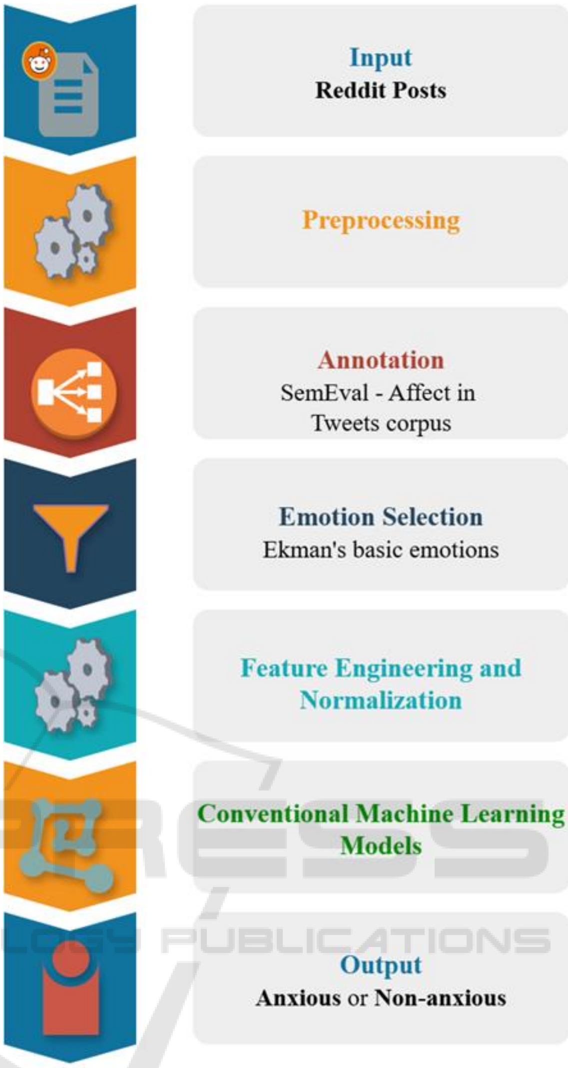

Figure 1: Procedural overview of detecting anxiety using emotion dynamics.

$$Min\ Emotion = \min \{E_{post,\ emotion_i}\} \quad (2)$$

$$Max\ Emotion = \max \{E_{post,\ emotion_i}\} \quad (3)$$

These aggregated scores offer a higher level of granularity and better reflect aspects of emotion dynamics compared to analysing raw post-level scores in isolation.

The normalized emotion gap is obtained as the ratio of the number of emotion gaps to the total emotion gap for a given user as per the following equation.

$$Normalized\ Gap_E = \frac{Total\ Gap_E}{Number\ of\ Gaps_E} \quad (4)$$

This measure provides insight into the average time between consecutive posts expressing a specific emotion, normalized to reflect the distribution and frequency of those emotional expressions for the user with the aim of incorporating temporal features in the analysis. The output, measured in seconds, offers a standardized assessment of the time between posts related to a specific emotion.

3.3 Models

Conventional machine learning models are utilized for the anxiety classification task. Our study employees five algorithms, focusing on decision tree, random forest, k-Nearest Neighbours (KNN), Linear Regression and Naïve Bayes models.

- Decision Tree: This supervised learning algorithm structures decisions in a tree-like format. The internal nodes represent feature-based decisions or tests, the edges indicate the outcomes of those decisions, and the leaf nodes hold the predicted values or class labels. Decision trees are known for their simplicity and interpretability.
- Random Forest (RF): An ensemble method that constructs multiple decision trees and aggregates their outputs to improve prediction accuracy. Each tree makes individual class predictions, and the final output is determined through majority voting or averaging the results from all trees.
- k-Nearest Neighbours (KNN): A straightforward, instance-based algorithm that does not require a training phase. Instead, KNN stores the entire dataset and makes predictions by identifying the k closest data points to a query point and using their majority class as the prediction.
- Naive Bayes: A probabilistic classifier based on Bayes' theorem, which assumes feature independence. It calculates the conditional probability of each class given the input features and assigns the class with the highest probability.
- Logistic Regression: A fundamental and interpretable algorithm used for binary classification tasks. It models the probability of a certain outcome using the sigmoid function and is most effective when the relationship between the dependent and independent variables is approximately linear.

3.4 Evaluation Metrics

The model’s performance is evaluated using several metrics, including the confusion matrix, precision, recall, F1 score, and accuracy. The confusion matrix classifies predictions into four categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). As shown in (Lee, Sohn, & Choi, 2019), precision measured the proportion of true positive predictions among all positive predictions, reflecting the model's ability to correctly identify anxiety cases. Recall evaluates the model’s capacity to correctly identify all actual anxiety cases, highlighting its sensitivity in (Ji, Ansari, Fu, Tiwari, & Cambria, 2021). The F1 score combines both precision and recall to provide a balanced measure of the model’s performance in (Aragon, Monroy, Gonzalez, Losada, & Montes, 2023). Finally, accuracy assesses how closely the model’s predictions aligned with the true values, indicating the overall correctness of the model in both anxiety and non-anxiety classes.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision + Recall}{TP + FN} \quad (8)$$

3.5 Feature Importance

Finally, we use SHAP (SHapley Additive exPlanations) to identify the feature importance of the best-performing model (Mosca, Tragianni, Gallagher, & Groh, 2022). SHAP offers a detailed and interpretable way to explain the output of machine learning models. This helps with further feature engineering, model interpretation, and building trust with the end users of the model. We used it to understand how each feature influences the best performing model's decision.

4 RESULTS AND DISCUSSION

In this study, we evaluate the performance of conventional machine learning models to classify emotional dynamics embedded in text data into two classes as anxious and non-anxious. The models tested include Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes. We used accuracy, precision, recall, and F1-score to critically evaluate the performance and results of these five models, examining their strengths and limitations in identifying anxiety instances and non-anxiety instances. Table 1 shows the results of each model.

4.1 Performance Evaluation

The Random Forest model outperformed the other models with an accuracy of 0.89. It demonstrated high precision for both classes: 0.73 for class 0 (non-anxious) and 0.92 for class 1 (anxious). The F1-scores of 0.68 for class 0 and 0.93 for class 1 indicate that Random Forest is highly capable of correctly classifying positive instances while maintaining a decent recall for class 0. This model's ability to handle data with a higher degree of reliability makes it a strong candidate for emotion dynamics analysis aligning with the findings of (Tate, et al., 2020).

The Decision Tree achieved an accuracy of 0.85, which is a commendable result. Its precision for class 0 was 0.60, and for class 1, it was 0.91, indicating that the model performed well in identifying anxious users but had a somewhat moderate performance for non-anxious users. Furthermore, the F1 score further confirms the model's strength in identifying anxious users with high accuracy but highlights its limitation in precision for non-anxious cases.

The F1-scores of 0.33 for class 0 and 0.88 for class 1 suggest that KNN is effective at identifying anxious cases, but it struggles significantly with identifying non-anxious cases. Logistic Regression demonstrated a significant discrepancy between the F1-scores for the two classes, indicating that it is highly effective at identifying positive instances but fails to capture negative instances adequately, showing signs of overfitting. Its recall for class 0 was notably low, which highlights its limitations in detecting non-anxious users. Naive Bayes exhibited the lowest overall performance, with an accuracy of 0.41. While it had a high recall for non-anxious posts, it struggled with precision and recall for anxious cases, resulting in an overall poor performance.

From the results, Random Forest emerged as the most accurate model for anxiety detection using emotion dynamics, providing high accuracy, precision, recall, and F1-scores across both classes. Decision Trees also demonstrated strong performance but with slightly lower precision for non-depressed posts. KNN, while effective at identifying depressive posts, showed significant weaknesses in detecting non-depressed posts. Logistic Regression was good at identifying depressive content but failed to detect non-depressed cases, as evidenced by its low recall for class 0. Finally, Naive Bayes, although it had a high recall for non-depressed posts, was unable to achieve reliable results for depressive cases, leading to its overall poor performance.

4.2 Feature Importance

The feature importance chart shown in Figure. 2 illustrates the contribution of emotion dynamics to the decision-making process in the best-performing model, which is Random Forest. The bar chart shows a significant correlation between the average sadness scores and anxiety disorder, which may reflect the persistent worry common among anxious individuals. The next important feature is the average joy score, which may demonstrate an inverse relationship with the anxious state, indicating a reduced positive effect. The higher importance of the average fear score may reflect fear and worry about future events, which is a common feature of anxiety. The normalized emotion gaps show intermediate feature importance, reflecting emotional swings, which could indicate emotional instability, a potential marker for anxiety. Furthermore, persistent anger can be an indicator of irritability in some individuals. Maximum emotion scores appear to be less related compared to average emotion scores and normalized emotion gaps, while minimum emotion scores show even lower feature importance, suggesting that anxiety is more influenced by elevated average negative emotions and reduced positive emotion averages.

Overall, in the decision-making process of the best-performing model (RF), both emotion intensity and emotion variability have a significant effect, with

Table 1: Evaluation results of machine learning models.

Model	Accuracy	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-Score (0)	F1-Score (1)
Decision Tree	0.85	0.6	0.91	0.62	0.9	0.61	0.9
Random Forest	0.89	0.73	0.92	0.64	0.94	0.68	0.93
KNN	0.80	0.46	0.84	0.26	0.93	0.33	0.88
Logistic Regression	0.82	0.75	0.82	0.07	0.99	0.13	0.9
Naive Bayes	0.41	0.23	0.92	0.9	0.3	0.37	0.45

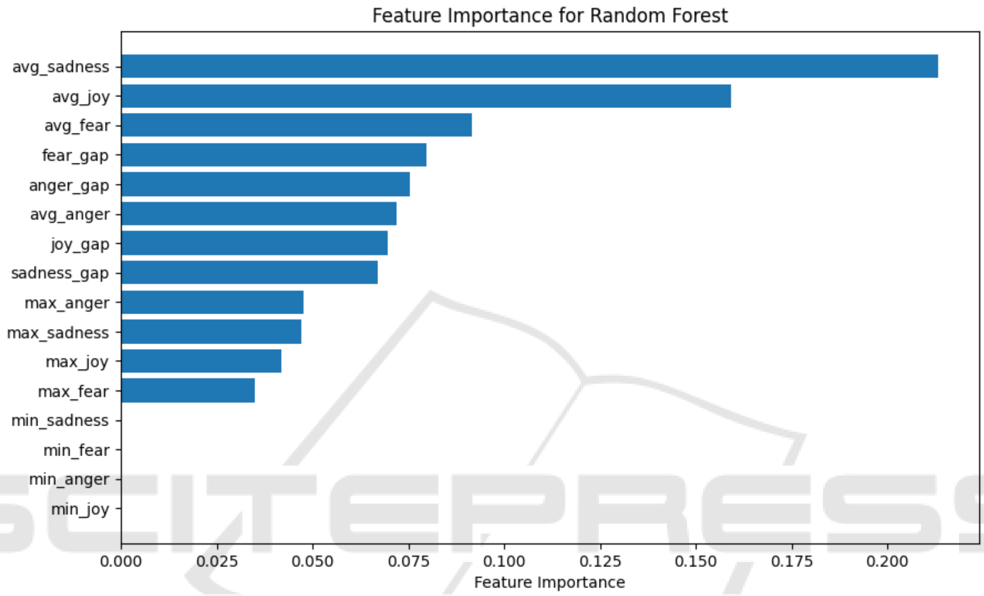

Figure 2: SHAP feature importance of the best performing model.

sadness and joy particularly influencing anxiety prediction, compared to other conventional machine learning models.

5 CONCLUSIONS

This study demonstrates the potential of using conventional machine learning models to analyse emotional dynamics in Reddit posts for anxiety detection. The random forest model outperformed the others, achieving 89% accuracy, with feature importance analysis highlighting the significance of average emotion scores and normalized emotion gaps. Specifically, emotions such as sadness and joy were found to play a crucial role in identifying anxiety. This emphasizes the value of emotion-dynamic-based analysis and conventional machine learning approaches in addressing complex mental health issues like anxiety detection and interpretation tasks. The study answers the two research questions: How can anxiety be classified based on basic emotional dynamics? and What is the impact of different emotional dynamics on anxious users?

However, several limitations should be acknowledged. First, the representativeness of the data may be limited, as Reddit users do not necessarily reflect the broader population of individuals with anxiety disorders. Additionally, the quality and context of user posts can vary, introducing noise and potentially impacting the accuracy of emotion detection. The study also focused on four basic emotions (sadness, joy, fear, and anger), limiting the scope of emotional dynamics analysed. Moreover, while the random forest model performed well, more advanced models such as deep learning techniques could potentially capture complex emotional patterns more effectively.

Future work should aim to address these limitations by incorporating a more diverse and representative dataset and exploring additional emotion categories to expand the emotion spectrum beyond the Ekman's basic four emotions. Enhanced feature engineering for average emotion scores and normalized emotion gaps could improve model interpretability and accuracy. Additionally, integrating temporal patterns and considering context within posts could provide more nuanced insights. Lastly, examining more sophisticated models and ensemble approaches, including deep learning, could improve performance and enhance the model's ability to detect signs of anxiety more reliably. These enhancements would further establish emotion-based machine learning models as a powerful tool for mental health detection and analysis.

REFERENCES

Anxiety disorders. (2024, April). Retrieved from National Institute of Mental Health: https://www.nimh.nih.gov/health/topics/anxiety-disorders

Aragon, M., Monroy, A. P., Gonzalez, L., Losada, D. E., & Montes, M. (2023). DisorBERT: A double domain adaptation model for detecting signs of mental disorders in social media. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 1, 15305-15318.

Baccianella, S., Esuli, A., Sebastiani, F., & others. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Lrec, 10, 2200-2204.

Bagroy, S., Kumaraguru, P., & De Choudhury, M. (2017). A social media based index of mental well-being in college campuses. Proceedings of the 2017 CHI Conference on Human factors in Computing Systems, 1634-1646.

Biagioni, R. (2016). The SenticNet sentiment lexicon: Exploring semantic richness in multi-word concepts. 4.

Buechel, S., & Hahn, U. (2022). Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. arXiv preprint arXiv:2205.01996.

Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. Proceedings of the SIGCHI conference on human factors in computing systems, 1909-1912.

Chatterjee, A., Narahari, K. N., Joshi, M., & Agrawal, P. (2019). SemEval-2019 task 3: EmoContext contextual emotion detection in text. Proceedings of the 13th international workshop on semantic evaluation, 39-48.

Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., & Goharian, N. (2018). SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. arXiv preprint arXiv:1806.05258.

De Choudhury, M., & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. Proceedings of the international AAAI conference on web and social media, 8, 71-80.

Ekman, P. (1999). Basic Emotions. In Handbook of Cognition and Emotion (pp. 45-60).

Guo, X., Sun, Y., & Vosoughi, S. (2021). Emotion-based modeling of mental disorders on social media. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 8-16.

Hasan, M., Rundensteiner, E., & Agu, E. (2019). Automatic emotion detection in text streams by analyzing twitter data. International Journal of Data Science and Analytics, 7, 35-51.

Hofmann, S. G., Moore, P. M., Gutner, C., & Weeks, J. W. (2012). Linguistic correlates of social anxiety disorder. Cognition & emotion, 26, 720-726.

Ji, S. a., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2021). Mentalbert: Publicly available pretrained language models for mental healthcare. arXiv preprint arXiv:2110.15621.

Kerz, E., Zanwar, S., Qiao, Y., & Wiechmann, D. (2023). Toward explainable AI (XAI) for mental health detection based on language behavior. Frontiers in psychiatry, 1219479.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of general psychiatry, 62, 617-627.

Ku, W. L., & Min, H. (2024). Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective Response Errors. Healthcare.

Lee, J., Sohn, D., & Choi, Y. S. (2019). A tool for spatio-temporal analysis of social anxiety with twitter data. Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 2120-2123.

Mohammad, S., Bravo-Marquez, F., Salameh, M., & Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. Proceedings of the 12th international workshop on semantic evaluation, 1-17.

Mosca, E. a., Tragianni, S., Gallagher, D., & Groh, G. (2022). SHAP-based explanation methods: a review for NLP interpretability. Proceedings of the 29th international conference on computational linguistics.

Nosov, A., Kuznetsova, Y., Stankevich, M., Smirnov, I., & Grigoriev, O. (2023). Modeling Seasonality of Emotional Tension in Social Media. Computers, 13(1), 3.

Qiao, W., Yan, Z., & Wang, X. (2024). When the clock chimes: The impact of on-the-hour effects on user anxiety content generation in social media platforms. Journal of Affective Disorders, 344, 69-78.

Saha, K., Yousuf, A., Boyd, R. L., Pennebaker, J. W., & De Choudhury, M. (2022). Social media discussions predict mental health consultations on college campuses. Scientific reports, 123.

Saifullah, S., Fauziah, Y., & Aribowo, A. S. (2021). Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. arXiv preprint arXiv:2101.06353.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. Journal of personality and social psychology, 66(2), 310.

Senanayake, A., Priyadarshana, Y., & Ranathunga, L. (2018). Nouns Speak: A Novel Approach for Noun Sentiment Scoring. 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC).

Shah, F. M., Reyadh, A. S., Shaafi, A. I., Ahmed, S., & Sithil, F. T. (2019). Emotion detection from tweets using AIT-2018 dataset. 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), 575-580.

Shen, J. H., & Rudzicz, F. (2017). Detecting anxiety through reddit. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality, 58-65.

Strapparava, C., & Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), 70-74.

Tate, A. E., McCabe, R. C., Larsson, H., Lundstrom, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. PloS one, e0230389.