User Mental Health Monitoring and Analyzing with NLP

1st Zmuda Mathias
School of Science and Technology (Computer Science)
IE University (of Aff.)
Madrid, Spain
mzmuda.ieu2022@student.ie.edu

2nd Etroy Baptiste
School of Science and Technology (Computer Science)
IE University (of Aff.)
Madrid, Spain
betroy.ieu2022@student.ie.edu

3rd Madkhana Wessam
School of Science and Technology (Computer Science)
IE University (of Aff.)
Madrid, Spain
wmadkhana.ieu2022@student.ie.edu

4th Bellens Victor
School of Science and Technology (Computer Science)
IE University (of Aff.)
Madrid, Spain
vbellens.ieu2022@student.ie.edu

***Abstract*—In modern day, the amount of text data that is made available every minute is so much more than we could imagine, and in this data is an extremely rich source of stories and clues from which we can infer the emotional well-being of a user. Analyzing emotional tones in written text provides a valuable opportunity to assess an individual's emotional well-being. Traditional Sentiment Analysis techniques only served to classify simple tones in text. However, recent advances in Natural Language Processing (NLP) allow for multiemotion detection. We look back to the astounding history and evolution of emotion detection and the introduction of NLP techniques and the only choice is to keep looking further. The current processes used for emotion classification with ground breaking methodologies are our pathway into successfully harnessing the most accurate form of user sentiment analysis there is, for the purpose of this paper we will be exploring and utilizing their potential applications in mental health monitoring.**

***Index Terms*—NLP, Sentiment Analysis, Neural Networks, Mental Health, Trends**

I. Literature Review

Sentiment analysis, also known as opinion mining, has been a critical research area in NLP, aiming to extract subjective information from textual data [1]. The field emerged in the early 2000s with foundational studies by Pang et al. (2002) and Turney (2002), which introduced statistical approaches to classify text into positive, negative, or neutral sentiments [1]. Early methods relied on lexicon-based approaches, where predefined word lists such as LIWC and NRC Emotion Lexicon were used for sentiment classification [1]. These approaches, while effective in capturing polarity, struggled with contextual nuances and figurative language such as sarcasm and irony.

Over time, researchers recognized the limitations of simple classification and sought to improve sentiment analysis using machine learning techniques. The introduction of Support Vector Machines (SVMs), Naïve Bayes classifiers, and Decision Trees allowed sentiment analysis to transition from rule-based systems to more adaptive, data-driven models [1]. However, these models were still limited in capturing the complexity of human sentiment, particularly in nuanced and domain-specific texts.

With the rise of deep learning, sentiment analysis saw a major transformation. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were utilized to capture sequential dependencies and extract semantic features from text [1]. Despite their success, these models struggled with long-range dependencies in text, leading to the development of transformer-based architectures such as BERT, RoBERTa, and XLNet, which significantly improved sentiment classification by leveraging self-attention mechanisms [1]. These advancements enabled multi-label classification and aspect-based sentiment analysis, making it possible to detect multiple emotions in a single text.

We want to leverage the newest and best models, while remembering the roots of Sentiment Analysis. We will do this by training our models with real-world mental health data on specific users and augment their functionality in order to birth a new way of analyzing user sentiment over time. Our project builds upon these innovations by expanding sentiment analysis beyond simple polarity classification. Unlike traditional approaches that classify text as merely positive or negative, our NLP-powered tool aims to detect a broad range of emotions, such as anxiety, joy, fear, and hope over time, creating an interactive map of a users emotions which our model will analyze and assess. By leveraging transformer-based models like RoBERTa and DistilBERT, and training them on GoEmotions and SemEval datasets, we aim to improve the accuracy and robustness of emotion detection.

The next paper we looked at was "**Improving the Generalizability of Text-Based Emotion Detection by Leveraging Transformers with Psycholinguistic Features**" [2] This paper showcases an interesting approach to enhancing emotion detection models' performance across datasets. In this, the authors have proposed combining transformer models(like BERT and RoBERTa) with Bidirectional Long Short-

Term Memory (BiLSTM) networks trained on a set of psycholinguistic features. This "hybrid" approach will essentially use both contextual representations from transformers and psychological nuances captured by BiLSTMs simultaneously, offering a stronger and more in-depth analysis of emotions in text.

A. Key Findings

- **Hybrid Model Architecture**: The integration of transformer embeddings with BiLSTM networks allows for effective contextual understanding while also incorporating psychological aspects of language [2].
- P**sycholinguistic Features**: The inclusion of features such as word concreteness, affective norms, and lexical diversity enhances the model's ability to detect the deeper emotional meaning behind text [2].
- **Evaluation**: The model is tested on benchmark datasets GoEmotions and ISEAR, showing superior performance in cross-domain emotion detection [2].

B. Relevance to Our Project

Our research aims to develop an NLP language model capable of analyzing text conversations to assess emotional states over time based on a users' textual activity. The findings from this paper align closely with our objectives, offering valuable insights into improving our model's accuracy and generalisability:

- **Model Architecture**: Incorporating a hybrid approach combining transformers with BiLSTM networks (trained on psycholinguistic features) will likely enhance emotion detection in our system.
- **Feature Engineering**: Leveraging psycholinguistic features will provide deeper insights into emotional nuances, enabling personalized emotional support for users.
- **Generalisability**: Ensuring our model performs well across various domains (e.g., text messages, Reddit, Twitter, and online discussions) is a priority. The methodologies outlined in [2] will help us achieve better out-of-domain generalization.

The applications that AI has in mental health research and analysis are countless, with one of the first ever chatbots, "Parry", giving us a taste of these endless possibilities with its paranoid and schizophrenic tendencies, mimicking a realistic mentally ill individual [3]. The advances mentioned in NLP with Sentiment Analysis have grasped a spot in the forefront of mental health technology innovation. The last paper we researched was "**Mental Health Assessment using Artificial Intelligence with Sentiment Analysis and NLP**" [4], which argues that AI-driven mental health tools can help address the shortage of mental health professionals by automating assessments, identifying risk factors, and providing early intervention. It discusses the use of machine learning and NLP to analyze speech and text patterns, making sentiment-based predictions of mental health conditions.

One of the primary advances in sentiment analysis for mental health discussed in this paper is the ability to detect mental health indicators through linguistic features. Traditional sentiment analysis often focuses on classifying text as positive, negative, or neutral, but this study explores a more refined approach using feature engineering techniques that allow for emotion-based sentiment classification. This shift is crucial for our project, as a simple sentiment classification is not enough to capture the complexity of mental health states. Instead, an emotion-based NLP approach, as described in this paper, enables multi-dimensional sentiment detection, which can distinguish between anxiety, stress, depression, and other affective states [4].

Furthermore, the research emphasizes how NLP techniques, such as sentiment lexicons, word embeddings, and contextual embeddings, play a role in understanding emotion in mental health-related text. The authors discuss how text classification methods combined with transformer-based models (e.g., BERT, RoBERTa) improve emotion detection accuracy, allowing a deeper understanding of human emotions beyond traditional polarity detection [4]. In our project, this means incorporating pretrained transformer models and custom sentiment lexicons that align with mental health discourse, ensuring that our system can detect distress signals more effectively.

Another major finding from the study is the use of large-scale mental health datasets for training AI models. The researchers exploited datasets such as the Mental Health Corpus (containing real-world conversations labeled with anxiety and depression indicators) and the Workplace Mental Health Survey, which helped improve model generalization [4]. Our project can incorporate similar mental health-related text corpora to train models in diverse sources such as Reddit, Twitter, and online mental health forums, ensuring that our model is robust across different styles and platforms of communication.

In addition, the study discusses the use of AI-driven chatbots for real-time mental health monitoring. The proposed system analyzes user interactions and provides emotion-based feedback, alerting users to potential mental health risks and, in extreme cases, recommending professional intervention [4]. Although our current project focuses on emotion detection, these findings provide a strong foundation for future expansions where we could develop interactive AI tools that offer real-time support based on emotional analysis.

II. Introduction

The vast majority of telltale signs that indicate a person's well-being are often observed through physical signals, such as body language, facial expressions, or daily habits [5]. However, when it comes to long-term emotional tracking, especially in digital spaces, these observables become largely difficult to track and analyze. Instead, we must rely on one of the strongest and most revealing indicators of mental state, speech and written communication patterns. The way individuals express themselves through language, whether consciously or subconsciously, can provide deep insight into their emotions, stress levels, and overall mental well-being [6]. For our model, we focus on analyzing online text data, particularly from avid social media users, where self-expression is frequent

and diverse. Digital communication platforms serve as a rich source of linguistic and emotional data, containing a vast spectrum of sentiments, tones, and emotional fluctuations. By tracking and interpreting these variations over time, we can construct a clearer picture of an individual's emotional state. Our goal was to develop a model capable of carefully and accurately processing vast amounts of textual information, particularly historical user activity. By harnessing rich user text datasets, performing deep sentiment analysis, and utilizing emotion classification models, our system will detect emotional patterns, identify peaks and troughs in sentiment, and assess long-term trends in emotional well-being. This will allow for a more structured understanding of mental health trends through digital footprints. Beyond passive analysis, our project aims to push sentiment detection a step further—potentially offering proactive insights or relevant healthcare recommendations in situations where severe emotional distress is detected. We wish for this to be a step towards autonomizing mental health research on social media.

III. Methodologies and Preliminary Analyses

A. Social Media Textual Dataset

The datasets selected for fine-tuning our sentiment analysis task comprise messages and posts generated by users across multiple social media platforms, including Twitter, Reddit, Facebook, Instagram, and various mental health forums. We intentionally opted for multiple data sources to enhance the generalisability and robustness of our sentiment analysis model. These datasets include anonymised textual data spanning various communication styles, ranging from brief posts and tweets to extensive and reflective messages, thereby providing comprehensive coverage of user-generated emotional expressions. Each textual entry is annotated with emotion labels consistent with the standards established by the GoEmotions and SemEval datasets. These annotations encompass a diverse array of emotional states, including but not limited to joy, sadness, fear, anxiety, anger, and neutral expressions.

In our preliminary exploratory data analysis, it was observed that the distribution of emotions is broadly balanced, though certain emotions, such as neutral and anxiety, appeared slightly more frequently. This diversity is beneficial for training robust models capable of generalizing well across different emotional contexts and textual styles encountered on social media platforms.

The data preprocessing involved several critical steps to ensure consistency and model compatibility. First, all data were aggregated into a unified dataset, ensuring standardization and uniform formatting. It was cleaned vigorously; We removed duplicate entries, defined a comprehensive text cleaning function that removed URLs using a regex pattern, mentions, hashtags, and special characters. Replace numbers with a placeholder or normalize them and convert all text to lowercase and trims extra spaces. Secondly, raw textual data underwent detailed preprocessing, which included:

- **Tokenization:** Segmented the text into individual tokens (words or subwords) using NLTK's `word_tokenize` function, with a fallback simple tokenizer when necessary.
- **Lemmatization and Stemming:** Tokens were processed using the WordNetLemmatizer to convert words to their base forms. On top of that, a stemming process using the Porter Stemmer was applied, thereby streamlining the vocabulary.
- **Stop-word Removal:** Removing common, low-information words(e.g., articles, prepositions) on a curated stopword list from NLTK to emphasize emotionally meaningful words with high semantic content.
- **Punctuation Normalization and Noise Removal:** Special characters, extraneous punctuation, URLs, mentions, and hashtags were removed using regular expressions. Also converted all characters to lowercase and eliminated redundant white spaces.
- **Numerical Normalization:** Numerical tokens were standardized by either replacing them with a placeholder or normalizing them.
- **Embedding Preparation:** Finally, we converted the cleaned and tokenized text into vector representations. TF-IDF vectorization was applied to capture term frequency and importance. Transformer-based embeddings were also generated to encode semantic relationships.

Following these preprocessing steps, texts were encoded into embeddings suitable for transformer-based neural networks.

The dataset was partitioned into three subsets: training (0.7), validation (0.15), and testing (0.15). This partitioning allows the model to learn robust emotional patterns, evaluate performance during training, and ensure unbiased performance validation on unseen data.

B. Models and Framework

For text classification, we selected transformer-based models RoBERTa and DistilBERT due to their superior ability to capture contextual nuances in emotional expressions. These models will be fine-tuned within the Kaggle computational environment, exploiting GPU resources for efficient training.

Our methodological framework involves an integrated pipeline:

- **Preprocessing and Embedding Generation:** As outlined above, preprocessing ensures the textual data are optimized for transformer-based models.
- **Model Training and Fine-Tuning:** Using annotated datasets, we fine-tune RoBERTa and DistilBERT to predict emotions accurately.
- **Emotion Probability Distribution:** Each model generates probability distributions for various emotional states. Missing emotion labels are handled with a default value (e.g. "unknown").
- **Aggregated Emotion Prediction:** Probabilities from both models are combined and averaged to identify the predominant emotional state per textual entry.

C. Graphical Representation and Interpretation

The results will be visualized through interactive graphical outputs, including detailed time-series plots that reveal emotional trajectories, highlighting critical emotional peaks, valleys, and gradual trends. These visualizations will provide intuitive and interpretable insights into emotional dynamics over time, facilitating comprehensive analysis and early identification of emotional fluctuations critical for proactive mental health interventions.

REFERENCES

[1] S.Poria, N. Majumder, R. Mihalcea, and D. Hazarika, "arxiv.org/pdf/2005.00357 16 nov 2020" Current challenges and New Directions in sentiment Analysis Research, https://arxiv.org/pdf/2005.00357 (accessed Feb. 18, 2025).

[2] S. Zanwar, D. Wiechmann, Y. Qiao, and E. Kerz, "ArXiv:2212.09465v1 [cs.CL] 19 dec 2022," Improving the Generalizability of Text-Based Emotion Detection by Leveraging Transformers with Psycholinguistic Features, https://arxiv.org/pdf/2212.09465.pdf (accessed Feb. 18, 2025).

[3] N. DR Jeevanandam, "Parry: The Pioneering Chatterbot of 1972," IndiaAI, https://indiaai.gov.in/article/parry-the-pioneering-chatterbot-of-1972 (accessed Feb. 18, 2025).

[4] S. Srivastava, S. Suchitra, V. Saraogi , and K. Arthi , "(PDF) mental health assessment using AI with sentiment analysis and NLP," ResearchGate, https://www.researchgate.net/publication/373906997_Mental_Health_Assessment_using_AI_with_Sentiment_Analysis_and_NLP (accessed Feb. 18, 2025).

[5] Hannah Owens, L. (2024, October 2). Are you sending the wrong signals? experts reveal the body language mistakes you might be making. Verywell Mind. https://www.verywellmind.com/how-to-read-mixed-signals-8719556

[6] Kasturi, B. (2016, May). Opinion detection, sentiment analysis and user attribute detection from online text data. Main page. https://alexandria.ucsb.edu/lib/ark