Pertanika J. Sci. & Technol. 33 (S4): 57 - 79 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Psychology-informed Natural Language Understanding: Integrating Personality and Emotion-aware Features for Comprehensive Sentiment Analysis and Depression Detection

Jing Jie Tan, Ban-Hoe Kwan*, Danny Wee-Kiat Ng and Yan Chai Hum

Department of Mechatronics and Biomedical Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, 43000, Kajang, Malaysia

ABSTRACT

This paper presents a novel approach to natural language understanding, integrating personality and emotion-aware features for sentiment analysis and depression detection. This research aims to enhance the performance of natural language understanding tasks, specifically sentiment analysis and depression detection, while also promoting explainability by including interpretable insights into psychological factors, such as emotion and personality, that influence these tasks. We refer to this additional feature as the psychology-informed module, alongside attention and transformer models. We achieved a significant improvement in accuracy using only the emotion feature: 3.4% for sentiment analysis on the IMDb dataset and 3.1% for depression detection on the SDCNL dataset. Similarly, using the personality feature led to a 2.5% improvement in sentiment analysis on the Polarity dataset and a 2.9% improvement in depression detection on the SDCNL dataset. On the other hand, the culmination of combining both psychological features achieves an accuracy of 0.8775 and 0.9053 for sentiment analysis on the Polarity and IMDb datasets, respectively. Additionally, notable results were obtained for depression detection, with accuracies of 0.8533 and 0.7177 on the Twitter (now known as X platform) and SDCNL datasets, respectively. These advancements enhance model accuracy and improve explainability, fostering versatile real-world applications. We thoroughly examined the factors, advantages, and limitations associated with this approach (psychology-informed module), providing a

ARTICLE INFO

Article history:
Received: 01 April 2024
Accepted: 10 April 2025
Published: 10 June 2025

DOI: https://doi.org/10.47836/pjst.33.S4.04

E-mail address:
tanjingjie@1utar.my (Jing Jie Tan)
kwanbh@utar.edu.my (Ban-Hoe Kwan)
ngwk@utar.edu.my (Danny Wee-Kiat Ng)
humyc@utar.edu.my (Yan Chai Hum)
*Corresponding author

e-ISSN: 2231-8526

comprehensive discussion within the scope of our study. The findings pave the way for future research to explore innovative techniques, further expanding the interdisciplinary impact of psychology-informed natural language understanding.

Keywords: Depression detection, emotion-aware recognition, machine learning, natural language processing, natural language understanding, personality-aware recognition, psychology-informed models, sentiment analysis

INTRODUCTION

In the contemporary digital era, Natural Language Understanding (NLU) is rapidly advancing and finding versatile applications in diverse domains such as recommendation systems, depression detection, sentiment analysis, and more. As we witness the growing influence of NLU in shaping technology-driven solutions, a compelling question arises: Can integrating psychological insights into text analysis enhance its performance?

Prior work has focused on enhancing sentiment classification performance in textual messages through integrating personality recognition (Tan et al., 2023). Building upon this foundation, the current study delves deeper into the potential synergy between Natural Language Understanding (NLU) and psychology. This research aims to improve model performance by incorporating psychological information (referred to as the psychology-informed module) into text analysis, contributing to more nuanced and enhanced outcomes. Subsequently, this research study examines the significance of improvements in this approach compared to the baseline using statistical testing.

By delving into the intricacies of human emotion, personality, and cognitive processes, we seek to elevate the capabilities of NLU, paving the way for more nuanced and contextually aware applications. This exploration is motivated by the belief that a deeper understanding of the psychological dimensions within language can unlock new possibilities for refining the accuracy and effectiveness of NLU systems in various real-world scenarios.

Incorporating a psychology-informed module enhances the performance of natural language understanding tasks. It promotes explainability by providing interpretable insights into the underlying psychological factors influencing sentiment analysis and depression detection. This emphasis on transparency contributes to a more trustworthy and comprehensible framework for decision-making in NLU applications. This approach holds unlimited potential; for instance, in real-time, harnessing emotions not only improves depression detection but also facilitates context-aware emotion labelling or user personality labelling, empowering the algorithm to deliver personalized recommendations to other users, thus achieving a dual benefit with a singular, versatile approach.

LITERATURE REVIEW

Psychology Informed Model

This research defines the psychology-informed model as a classical model enhanced with a psychology-informed module, incorporating features derived from personality recognition and emotion recognition.

Personality Recognition

Various personality theories propose distinct dimensions (Cervone & Pervin, 2022), including the Eysenck Personality Questionnaire (EPQ) with the Big-3 model (Eysenck & Eysenck, 1975), the Myers-Briggs type indicator (MBTI) represents the Big-4 model, and the widely accepted Big-5 model encompasses "openness, conscientiousness, extraversion, agreeableness, and neuroticism" (OCEAN) (Goldberg, 1993). Additionally, the Big-6 model, incorporating "honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness" (HEXACO) (Ashton et al., 2004) provides another perspective on personality dimensions. According to Moreno et al. (2021), personality traits of an individual are shown in his or her written text.

Mairesse et al. (2007) made a significant contribution by conducting a state-of-the-art research focused on psycholinguistic features. Their study on the Essays corpus employed correlational analysis to identify key features influencing personality classification. In addition, Sun et al. (2019) justified topic word extraction through lexicons and word2vec for measuring personal traits in specific aspects using user-generated text. These techniques demonstrate significant diversity among individuals in affect and social interaction, revealing correlations with personality traits. In addition, transformer embedding includes Bidirectional Encoder Representations from Transformers (BERT) with SenticNet (a psychological lexicon), which shows improvement in personality recognition (Ren et al., 2021). These show the potential of personality recognition to improve the natural language understanding model.

Emotion Recognition

Emotions serve as immediate indicators of our psychological state. Circumplex Model of Affect is a widely acknowledged framework that categorizes emotions within a two-dimensional space defined by valence (pleasant-unpleasant) and arousal (low-high), offering a structured representation of diverse emotional experiences and their relationships (Posner et al., 2005). The seven common emotions are joy, anger, fear, surprise, sadness, disgust, and contempt. However, we can further fine-grain emotions such as admiration, love, and others. Hence, several datasets such as the ISEAR dataset (Scherer & Wallbott, 1994), the DENS dataset (C. Liu et al., 2019), and the GoEmotions dataset (Demszky et al., 2020) featured varying numbers of classes, commonly utilized in emotion machine learning research.

Notably, for concise text, long short-term memory (LSTM) demonstrates superior performance with an accuracy of 97.50%, surpassing Linear Support Vector Classifier, which achieves only 89% (Alfarizi et al., 2022). Subsequent advancements involve transformer embeddings, revealing enhanced results with accuracy scores of 0.7431, 0.7299, 0.7009, and 0.6693 for its variations, namely RoBERTa, XLNet, BERT, and DistilBERT, respectively (Adoma et al., 2020). Therefore, this study delves into evaluating the efficacy of extracting emotional information features from transformer models as opposed to traditional statistical models.

Natural Language Understanding

To validate our hypothesis, which incorporates insights from psychological information, specifically personality and emotion, we focus on two natural language understanding tasks: sentiment analysis and depression detection.

Sentiment Analysis

In today's digital age, understanding and effectively classifying sentiments expressed in textual messages has become crucial for a wide range of applications. Sentiment analysis plays a pivotal role in enhancing various aspects of communication, business, and social interactions. (Yang et al., 2020; Yarkoni, 2010; Lin et al., 2017). It involves binary classification tasks, commonly applied to datasets that include movie reviews from movie and shopping platforms such as IMDb (Maas et al., 2011a) and Amazon Review (Keung et al., 2020).

Researchers explored the combination of lexicon-based methods and neural networks for sentiment analysis, achieving comparable accuracy of 90% when employing a recurrent neural network (RNN) and their variant, long short-term memory (LSTM), on the IMDb dataset (Qaisar, 2020; Shaukat et al., 2020). Meanwhile, on the Amazon Review dataset, a TF-IDF approach with logistic regression yielded an accuracy of 0.9; similarly, a BERT model achieved a comparable level of accuracy (Durairaj & Chinnalagu, 2021; Rajat et al., 2021). The observed identical accuracies prompt a consideration for potentially adopting a more efficient or lightweight approach to achieve similar results in sentiment analysis tasks.

Depression Detection

Depression detection models are significant as a screening tool to facilitate early treatment (Souza Filho et al., 2021). According to Havige et al. (2019), there is a strong positive correlation between a written message and the risk or the degree of depression in an individual. Hence, research has shown the possibility of detecting a person's depression based on a text message that he or she posted or commented on. A study on postpartum depression also supported that the linguistic style in a message's content is a major indicator that is able to predict whether a person has depression or not (De Choudhury et al., 2014). There are widely used datasets, such as the Reddit and SDCNL datasets, which were collected through web scraping from users' self-reports on social media platforms (Haque et al., 2021; Yates et al., 2017).

Choudhury et al. (2016) proved the positive correlation between emotion and linguistic style labelled by LIWC and depression. Using a support vector machine classifier, the overall depression detection accuracy is 0.68. According to Kamal et al. (2018), the highest accuracy achieved in depression classification is 0.73 using the decision tree algorithm.

Tadesse et al. (2019) explored various combinations of lexicon features and models, finding that the combination of linguistic inquiry and word count (LIWC) + Latent Dirichlet Allocation (LDA) + bigram with a multi-layer perceptron (MLP) classifier achieved a noteworthy 91% accuracy. This underscores the significance of incorporating psychological insights, as lexicons carry nuanced meanings. Additionally, Figuerêdo et al. (2022) conducted further experiments, revealing that the semantic mapping of emoticons resulted in a 0.05 improvement in F1 score, highlighting the potential impact of emotion in text on depression detection.

METHODS

We aim to enhance natural language understanding (NLU), specifically in sentiment and depression classification, by integrating a classical NLU model with a psychology-informed module, incorporating personality recognition and/or emotion recognition results. This integration provides additional insights into the psychological status of the classifier. Our experiments encompass various combinations, including utilizing psychology-informed models and datasets, to enhance the overall performance of the classification tasks.

As illustrated in Figure 1, we commence the process by training a psychology-aware model to achieve this. This model will then generate psychological features for subsequent NLU model training. Finally, we employ this model for evaluation.

The related resources (algorithm repository and split datasets for reproducibility) are all in one place: https://research.jingjietan.com/?q=PSYCHONLU.

Datasets

Table 1 provides an overview of the datasets utilized in this research. Two widely recognized and relevant datasets were selected for each model's task domain. This choice allows for a comparative analysis that may offer additional evidence to support the research hypothesis. Selection criteria focused on datasets with an average size of approximately 1,000 to 100,000 records—ideal for training and analysis without being excessively large.

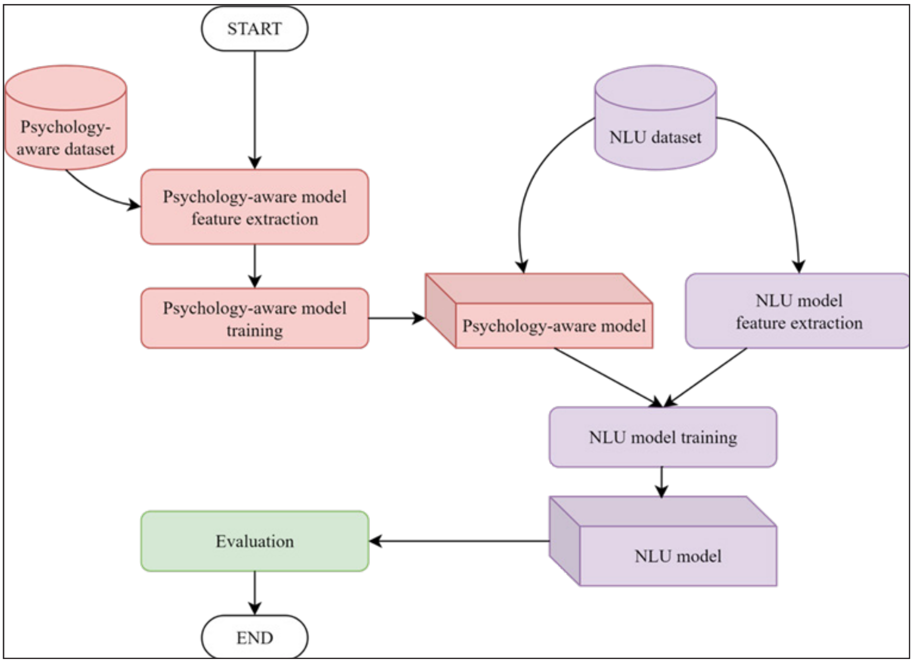

Figure 1. The block diagram of the proposed model architecture

Table 1
The datasets used in the experiments

Type	Model Task	Dataset	Description & Splitting
Psycho	Personality Recognition	Kaggle-MBTI-PersonalityCafe (Mitchell, 2017)	The dataset was gathered from Reddit and comprises documents labelled with MBTI (Myers-Briggs Type Indicator) types, which include 16 specific types or four binary labels. https://huggingface.co/datasets/jingjietan/kaggle-mbti (Tan, 2024b)
		Reddit-MBTI (Deimann et al., 2023)	The dataset, available upon request, was gathered from Reddit and consists of documents labelled with MBTI types, encompassing 16 distinct personality types or four binary category labels.
	Emotion Recognition	GoEmotions (Demszky et al., 2020)	Gathered from Reddit, this dataset includes text and categorizes it into 27 distinct emotions and a neutral category. https://huggingface.co/datasets/willcb/go-emotion (Demszky et al., 2020)
		CombinedDataset (Hartmann, 2022)	Merged from six open datasets, this combined dataset was refined by removing certain classes, resulting in a final set of seven classes. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base (Hartmann, 2022)

Table 1 *(continue)*

Type	Model Task	Dataset	Description & Splitting
NLU	Sentiment Analysis	Polarity (v2.0) (Pang & Lee, 2004)	Comprising 1,000 movie review comments, this dataset is labelled with positive or negative sentiments. https://huggingface.co/datasets/jingjietan/polarity-sentiment (Tan, 2024c)
		IMDb (Maas et al., 2011a)	Composed of 50,000 comments on movie reviews, expressed in positive or negative terms. https://huggingface.co/datasets/jingjietan/imdb-sentiment (Tan, 2024a)
	Depression Detection	Twitter (now known as X) (Shinde, 2022)	Derived from Twitter posts, this dataset consists of 20,000 entries labelled as either indicating depression or not. https://huggingface.co/datasets/jingjietan/twitter-depression (Tan, 2024e)
		SDCNL (Haque et al., 2021)	Gathered from 1,895 Reddit posts, this dataset is labelled based on the presence of suicide intent. https://huggingface.co/datasets/jingjietan/sdcnl-suicide (Tan, 2024d)

Psychology-informed Modelling

We utilize the model parameters from the paper or train the model using the prepared dataset.

Personality Recognition

We employ the Kaggle-MBTI-PersonalityCafe dataset to train a **TF-IDF** model. TF-IDF (term frequency-inverse document frequency) is a numerical representation of the importance of a word in a document within a collection of documents, emphasizing terms that are frequent in a specific document but rare across the entire collection (Aizawa, 2003). The four binary features: I/E, S/N, T/F, or P/J represent distinct dimensions of the Myers-Briggs type indicator (MBTI), capturing preferences in terms of introversion/extraversion (I/E), sensing/intuition (S/I), thinking/feeling (T/F), and judging/perceiving(J/P), respectively (Sonmezoz et al., 2020). The formula for TF-IDF is represented by Equation 1, where tf_s represent the frequency of the term t in document s, while N is the number of documents in d, and df_t is the number of documents containing the term t.

$$TF-IDF(t,s) = TF(t,s) \times IDF(t) = tf_s \times ln\frac{N}{df_t} \quad [1]$$

Here, we have prepared four binary classifiers, each producing a label for an MBTI dimension. The data pre-processing involves text cleaning to remove irrelevant links and symbols, followed by tokenization, as in Algorithm I.

Algorithm I: Data Pre-processing

Input: Dataset *D*
Regex for link *R_LINK* ← *https?://[^\s<>"]+|www\.[^\s<>"]+*”
Regex for symbol *R_SYMBOL* ← “*[^0-9a-z]*”

Output Dataset *D*

 # Assume N dimension is available in this dataset
GET $texts$ **FROM** D_d
 # Loop all the text
FOR $text \in texts$ **DO**:
    $text \leftarrow Lower(text)$
    $text \leftarrow regex.sub(R_{LINK}, \ , text).sub(R_{SYMBOL}, , text)$
    $text \leftarrow text.split()$
ENDDO
RETURN *D*

We process both the training set and test set using the same tokenizer. Subsequently, we compute the TF-IDF for each word in the dataset, resulting in floating-point values. To maintain consistent input lengths for the neural network model, we set the maximum length to 5000. The Algorithm II outlines the training process for the multilayer perceptron model.

Algorithm II: Multilayer Perceptron Model Training

Input: Dataset *D*
Model, M
Hyperparameter $(epochs, \theta \ldots)$

Output Model, M

FOR $(text, label) \in (D_{text}, D_{label})$ **DO:**
  SET $text, label$ **TO** $trainloader$
  FOR e **IN** $epochs$
    FOR $inputs, targets$ **IN** $trainloader$
      $outs \leftarrow FeedForward(M, inputs)$
      $loss \leftarrow Loss(outs, targets)$
      UPDATE M, $M \leftarrow M - \theta \frac{\partial L(\hat{y}, y)}{\partial M}$ *# Back propagate with learning rate* θ*:*
    ENDDO
  ENDDO
ENDDO
RETURN M

We prepared a model utilizing the parameters that were pre-trained from ALBERT (Robert Deimann et al., 2023) using the Reddit-MBTI Dataset. ALBERT (A Lite BERT) addresses challenges in scaling up model size during pretraining, offering parameter-reduction techniques to enhance memory efficiency and training speed (Lan et al., 2019). This model is a multiclass model that outputs the probability for each MBTI class.

The overall process begins with data cleaning and removing HTML tags and irrelevant information, as outlined in Algorithm I. Subsequently, we utilize the tokenizer from the respective transformer to tokenize the text. Finally, we feed the tokenized values into the transformer model, including the output layer, to obtain the informed features.

There are two ways to incorporate psychology-informed features. First, with ALBERT-first, we use the highest probability as the recognition output, such as INFP, resulting in four binary features with values of 0 or 1. Second, we included all classes as features with the **ALBERT-list**, resulting in 16 features, each corresponding to an MBTI type, such as INFP.

Emotion Recognition

Likewise, we employ Algorithm I and Algorithm II in personality recognition to train a **TF-IDF** model on the GoEmotions dataset. We also included three additional models, producing psychology-informed results using parameters from transformer models, as detailed in Algorithm III.

Algorithm III: Model Inference from Pretrained or Referenced Model

Input: Dataset D_{in}

Output Dataset D_{out}

MAKE $tokenizer$ **FROM** $vocabulary$

CONSTRUCT $model$ **USING** $pretrained\ or\ referenced\ parameters$

$D_{tokenise} \leftarrow tokenizer(D_{in})$

$D_{out} \leftarrow tokenizer(D_{tokenize})$

RETURN D_{out}

Firstly, we employ the **BERT** model parameters fine-tuned by Raw (2021) on the CombinedDataset, yielding six emotion feature labels. BERT is a transformer model developed by Devlin et al. (2018) that represents a cornerstone in natural language processing due to its versatility and effectiveness in capturing contextual information in text.

Next, we leverage RoBERTa, an optimized transformer model utilizing dynamic masking during pretraining. This alteration enhances the model's capture of contextual information, improving performance across various natural language processing tasks (Y.

Liu et al., 2019). This model has two variations: the distilled (light) version, **DistilRoBERTa** and the standard **RoBERTa**. For the former, reconstructing with parameters from Hartmann (2022) results in a 7-class multiclass output. Subsequently, utilizing Lowe's (2023) model parameters, we obtain a 28-multilabel model.

Natural Language Understanding

Despite the availability of various feature extraction methods and models for NLU tasks, we opted for TF-IDF and a multilayer perceptron to ensure a fair comparison. We utilize the psychology-informed models mentioned above to generate features for NLU, specifically for **sentiment analysis** and **depression detection**. These features serve as additional support alongside the traditional data cleaning and TF-IDF feature extraction processes outlined in Algorithm I. The resulting 5000 features are concatenated with the psychology features and fed into a neural network, as illustrated in Figure 2.

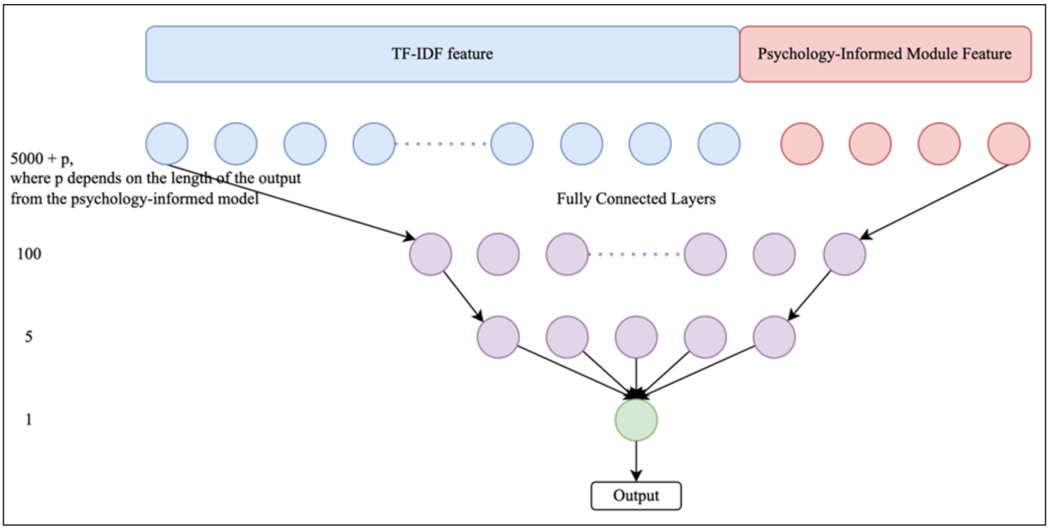

Figure 2. Illustration of combining psychology-informed model features with TF-IDF for multilayer perceptron (MLP) training by concatenation: e.g., the four personality outputs (represented by red circles) are included alongside TF-IDF features (represented by blue circles) for sentiment analysis

However, the concatenation process does impact the total number of parameters. Nevertheless, we have found that its impact is negligible, as the additional features contribute minimally. The mathematical proof is provided below:

The number of parameters between layers is = $n^{[l]} \times n^{[l+1]} + n^{[l+1]}$, where $n^{[0]}$ indicates the number of neurons in layer l.

Hence, the equation is to ensure an equal number of parameters for the normal model and the psychology-informed model, as shown in Equation 2:

$$\left(n_{norm}^{[0]} \times n_{norm}^{[1]} + n_{norm}^{[1]}\right) + \left(n_{norm}^{[1]} \times n_{norm}^{[2]} + n_{norm}^{[2]}\right) = \left(n_{psycho}^{[0]} \times n_{psycho}^{[1]} + n_{psycho}^{[1]}\right) + \left(n_{psycho}^{[1]} \times n_{psycho}^{[2]} + n_{psycho}^{[2]}\right) \quad [2]$$

Then, by arranging,

$$n_{psycho}^{[1]} = \frac{\left(n_{norm}^{[0]} \times n_{norm}^{[1]} + n_{norm}^{[1]}\right) + \left(n_{norm}^{[1]} \times n_{norm}^{[2]} + n_{norm}^{[2]}\right) - n_{psycho}^{[2]}}{1 + n_{psycho}^{[2]} + n_{psycho}^{[0]}} \quad [3]$$

As shown in Figure 2, substituting for the normal case values, where the input features, $n_{norm}^{[0]}$ is set to 5000, and the first hidden layer, $n_{norm}^{[1]}$, is set to 100. For input features of the psychology-informed model, it is the summation of TF-IDF features and psychology-informed features, denoted as $n_{psycho}^{[0]}$, equal to 5000 + 28, (28 being the highest number of features throughout the experiments. Lastly, the second hidden layer is the same for both models, $n_{norm}^{[2]} = n_{psycho}^{[2]} = 5$.

$$n_{psycho}^{[1]} = \frac{(5000 \times 100 + 100) + (100 \times 5 + 5) - 5}{1 + 5 + 5028} = 99.44 \approx 99 \quad [4]$$

From Equation 3, the suggested number of neurons is rounded to 99 (the normal setting is 100, which is not significantly different). We conducted experiments to further examine this setting and observed no noteworthy impact.

Evaluation

Lastly, the experiment was designed with a data split of 80% for training and 20% for testing, where the training set was further divided into an 80:20 ratio for the training set and the validation set. The test set remains invisible during training. The evaluation metric used is accuracy, formulated as Equation 5, where TP represents True Positive, TN is True Negative, FP is a False Positive, and FN is False Negative.

$$Accuracy = \frac{number\ of\ correct\ prediction}{number\ of\ sample} = \frac{TP + TN}{TP + TN + FP + FN} \quad [5]$$

Moving on, to demonstrate the significance of the proposed psychology-informed module, we employ the McNemar test—a statistical method for analyzing paired categorical data (Smith & Ruxton, 2020). This test helps assess whether the intervention (the proposed psychology-informed module) leads to a statistically significant change in the model's classification outcomes (e.g., correct vs. incorrect predictions) compared to the performance with and without the module.

The null hypothesis (H_0) and alternative hypothesis (H_1) for the McNemar test are formulated as follows:

- **Null Hypothesis ($\boldsymbol{H_0}$)**: There is no significant difference in the outcomes of the model with and without the psychology-informed module.
- **Alternative Hypothesis ($\boldsymbol{H_1}$)**: There is a significant difference in the outcomes of the model with and without the psychology-informed module, indicating that the module leads to a significant change in the outcomes.

The McNemar test specifically examines the count of discordant pairs in the data, cases where outcomes differ between the two conditions. The test statistic, χ^2, is calculated as shown in Equation 6:

$$\chi^2 = \frac{(b-c)^2}{b+c} \qquad [6]$$

where:

- b is the number of cases where the model's prediction was correct before the addition of the psychology-informed module but incorrect after,
- c is the number of cases in which the prediction was incorrect before the addition of the psychology-informed module, but was correct after.

To evaluate statistical significance, we calculate the p-value using the cumulative distribution function (CDF) of the chi-squared distribution with degrees of freedom (df) set to 1, using Equation 7:

$$p = 1 - CDF(\chi, df = 1) \qquad [7]$$

We then compare the p-value to a common significance level, $\boldsymbol{\alpha = 0.05}$. If the p-value is less than 0.05, we reject the null hypothesis, indicating that the psychology-informed module has a statistically significant effect on the model's outcomes. Conversely, suppose the p-value is greater than or equal to 0.05. In that case, we fail to reject the null hypothesis, suggesting insufficient evidence to conclude that the module significantly affects model performance.

RESULTS AND DISCUSSION

The results are presented in a table, highlighting the improvements achieved by the proposed psychology-informed module compared to the regular method. Improvements with a highly significant p-value ($p \leq 0.01$) are indicated with an asterisk in **bold***. Additionally, improvements with a p-value between 0.01 and 0.05 ($0.01 < p < 0.05$) are also shown in **bold**. Any outcomes that do not show a significant improvement (where the null hypothesis is not rejected) are displayed in regular font, indicating no statistically significant difference between the model with and without the psychology-informed module.

Sentiment Analysis

As evident in Tables 2 and 3, there is a notable improvement in sentiment classification achieved through applying psychology-informed modules. Nevertheless, the Polarity dataset is smaller (~25× smaller) compared to the 1MDb dataset, which may lead to certain models failing to meet the statistical significance threshold for the hypothesis.

Table 2
Sentiment analysis results with personality-aware techniques

Dataset	Personality-aware Model	Accuracy	Improvement
Polarity	Regular (None)	0.8375	-
	TF-IDF	0.8625	**0.0250**
	ALBERT (list)	0.8525	0.0150
	ALBERT (first)	0.8600	**0.0225**
IMDb	Regular (None)	0.8710	-
	TF-IDF	0.8854	**0.0144***
	ALBERT (list)	0.8740	**0.0030**
	ALBERT (first)	0.8845	**0.0135***

Table 3
Sentiment analysis results with emotion-aware techniques

Dataset	Emotion-aware Model	Accuracy	Improvement
Polarity	Regular (None)	0.8375	-
	TF-IDF	0.8600	**0.0225**
	BERT	0.8600	0.0225
	RoBERTa	0.8600	0.0225
	DistilRoBERTa	0.8650	**0.0275**
IMDb	Regular (None)	0.8710	-
	TF-IDF	0.8923	**0.0213***
	BERT	0.8748	**0.0038***
	RoBERTa	0.9051	**0.0341***
	DistilRoBERTa	0.8929	**0.0219***

However, some psychology-informed modules still yield p-values less than 0.05, indicating a statistically significant impact, albeit at a lower confidence level. This is sufficient to demonstrate the effectiveness of the module within the smaller dataset. In contrast, the 1MDb dataset shows stronger significance, with most models achieving p-values well below 0.05, further validating the module's effectiveness in a larger dataset.

Moving on, particularly noteworthy is the observation that emotion-aware models exhibit more substantial enhancements compared to their personality-aware counterparts. Specifically, the RoBERTa model, when applied to the IMDb dataset, demonstrates a notable improvement of 3.4%. This significant boost can be attributed to its capacity to tabulate 28 classes of emotion in probability, providing the model with valuable information for determining the overall sentiment.

In contrast, the performance of the BERT model in the IMDb dataset, which outputs six classes of features, shows the lowest accuracy. A similar limitation is observed with DistilRoBERTa, which outputs seven classes, merely 1% away from the more comprehensive RoBERTa model. This constraint in class representation underscores the significance of having a more extensive set of classes, influencing the model's capacity to discern nuanced sentiments in complex textual data.

Next, the RoBERTa and DistilRoBERTa emotion-aware models do not exhibit a significant difference in accuracy on the Polarity dataset. Upon further examination of both datasets, it is evident that the text in the Polarity dataset is relatively shorter (~4x) than that in the IMDb dataset. This difference leads to the conclusion that the information from the psychology-informed model does not have a substantial impact, as it is likely insufficient for the model to conduct a deep analysis of emotion. This explanation clarifies why the DistilRoBERTa model, which outputs seven features, performs similarly to the RoBERTa model, which outputs 28 features in the Polarity dataset.

Additionally, it is essential to note that transformer-based psychology-informed models do not manifest a significant improvement over the TF-IDF method. This observation underscores the notion that a sophisticated psychology-informed model is unnecessary to achieve enhanced accuracy. Nevertheless, using RoBERTa and DistilRoBERTa typically yields better results than the BERT model in psychology-informed applications, as they benefit from enhanced training strategies, including larger datasets and extended training periods without the Next Sentence Prediction task (Y. Liu et al., 2019).

Depression Detection

As evident in Tables 4 and 5, there is a notable improvement in depression detection achieved through the application of psychology-informed modules. Nevertheless, the Twitter dataset is approximately 10 times larger compared to the SDCNL dataset. This larger sample size enhances the statistical power of the analysis, making it more likely to detect significant differences. As a result, the models on the Twitter dataset exhibit stronger statistical significance, with p-values well below the threshold of 0.05, indicating a clear and meaningful impact of the psychology-informed module. In contrast, the smaller SDCNL dataset may not provide enough data to achieve similar levels of statistical significance, which might result in less pronounced findings.

Table 4
Depression detection results with personality-aware techniques

Dataset	**Personality-aware Model**	**Accuracy**	**Improvement**
Twitter	Regular (None)	0.8485	-
	TF-IDF	0.8523	**0.0037**
	ALBERT (list)	0.8515	**0.0030**
	ALBERT (first)	0.8525	**0.0040***
SDCNL	Regular (None)	0.6834	-
	TF-IDF	0.7071	**0.0237**
	ALBERT (list)	0.6939	0.0106
	ALBERT (first)	0.7124	**0.0290**

Table 5
Depression detection results with emotion-aware techniques

Dataset	**Emotion-aware Model**	**Accuracy**	**Improvement**
Twitter	Regular (None)	0.8485	-
	TF-IDF	0.8555	**0.0070***
	BERT	0.8505	**0.0020**
	RoBERTa	0.8540	**0.0055***
	DistilRoBERTa	0.8560	**0.0075***
SDCNL	Regular (None)	0.6834	-
	TF-IDF	0.7098	**0.0264**
	BERT	0.6939	0.0106
	RoBERTa	0.7018	0.0185
	DistilRoBERTa	0.7150	**0.0317**

In both datasets, it is generally observed that ALBERT (list) consistently performs worse than ALBERT (first). This discrepancy can be attributed to the distinction between the personality-aware and emotion-aware models. The former necessitates using only the highest possible value, as individuals typically belong to one personality type. In contrast, an emotion-aware model acknowledges the potential coexistence of multiple emotions simultaneously. For instance, in scenarios like a birthday party context, emotions such as surprise and joy can co-occur.

Additionally, the BERT model shows little improvement compared to the DistilRoBERTa model, which outputs 28 emotion probabilities. This observation underscores, once again, the importance of having a sufficient number of generated features from psychology-informed modules.

Personality Recognition

In addition, personality recognition is considered part of Natural Language Understanding (NLU), with emotion as the psychology-informed module. This approach is based on the premise that emotions expressed in a given context can potentially reflect aspects of an individual's personality. Table 6 illustrates the potential of leveraging emotion for MBTI classification, with results typically showing significant improvements in certain personality dimensions, particularly in the **P/J** (perceiving/judging) and **T/F** (thinking/feeling) dimensions. These dimensions are more likely to be influenced by the emotion-based insights, highlighting the value of incorporating emotional context into personality recognition models.

Table 6
Personality recognition results with emotion-aware techniques

Personality Dimension	Emotion-aware Model	Accuracy	Improvement
	Regular (None)	0.8818	-
S/N	RoBERTa	0.8835	0.0017
	DistilRoBERTa	0.8888	**0.0068**
	Regular (None)	0.7608	-
P/J	RoBERTa	0.7740	**0.0133***
	DistilRoBERTa	0.7671	**0.0063**
	Regular (None)	0.8249	-
I/E	RoBERTa	0.8251	0.0002
	DistilRoBERTa	0.8329	**0.0081**
	Regular (None)	0.8104	-
T/F	RoBERTa	0.8202	**0.0098***
	DistilRoBERTa	0.8294	**0.0190***

However, it is crucial to note that further experiments are needed to better understand why certain personality traits contribute more strongly to emotional expression, especially when considering factors such as dataset imbalance and the data collection method. For example, individuals with certain personality types may be less inclined to post on social media, resulting in a lower representation of these personality types in the dataset. This imbalance can influence the model’s ability to generalize and may lead to skewed results, underscoring the need for additional research to account for these factors.

Benchmark

To obtain the final evaluation of the model, we combine two psychology-informed features: personality (ALBERT (first)) and emotion (RoBERTa). The improvement is tabulated based on the highest accuracy achieved by a single model, as shown in Equation 8.

$$improvement = accuracy_{combine} - max\left(accuracy_{regular}\ , accuracy_{personality}\ , accuracy_{emotion}\ \right) \quad [8]$$

Table 7 shows that the improvement is relatively limited when combining personality and emotion modules, lacking significant significance compared to the individual (only personality or emotion module) models. Nonetheless, the Polarity dataset shows a noteworthy improvement of 1.7% (p-value = 0.10), which, while marginally above the typical significance threshold, is considered meaningful given the dataset's limited size. This finding suggests potential in the psychology-informed models for short text lengths. It highlights the need for further investigation in future work to better understand the model's performance on brief text data. Nevertheless, all observed improvements for psychology-informed NLU models are statistically significant compared to the regular approach models.

Table 7
Proposed approach: Psychology-informed NLU (Personality-aware (ALBERT) + Emotion-aware (RoBERTa))

Dataset	**Reference Accuracy**			**Accuracy (Personality-aware (ALBERT) + Emotion-aware (RoBERTa))**	**Improvement**	
	Regular (None)	**Personality (ALBERT (first))**	**Emotion (RoBERTa)**		**Compared to Regular (None)**	**Compared to Single Feature (Personality or Emotion) (Best)**
Polarity	0.8375	0.8600	0.8600	0.8775	**0.0400**	0.0175
IMDb	0.8710	0.8845	0.9051	0.9053	**0.0343***	0.0002
Twitter	0.8485	0.8525	0.8540	0.8553	**0.0068***	0.0013
SDCNL	0.6834	0.7124	0.7018	0.7177	**0.0343**	0.0053

Last but not least, we tabulate results from other researchers in Table 8 for comparative analysis. This comparison reaffirms that psychology-informed models demonstrate the capability to enhance accuracy, even when compared to heavy transformer models.

Even with a simple TF-IDF-based psychology-informed model, such as emotion-aware recognition for sentiment analysis on the IMDB dataset, our approach achieves an accuracy of 0.8923 (refer to Table 3), surpassing the 0.8905 achieved by the benchmark ALBERT model (refer to Table 8). Notably, our approach maintains superiority with significantly lower computational costs due to its lightweight design. This underscores the potential effectiveness of integrating psychological insights into the NLU framework for improved performance.

Table 8
Benchmark comparison of proposed approach: Psychology-informed NLU (Personality-aware (ALBERT) + Emotion-aware (RoBERTa))

Dataset	Approach	Accuracy
Polarity	**Proposed Psychology-Informed Approach**	**0.8775**
	Support Vector Machine (SVM) + Information Gain (Maulana et al., 2020)	0.8565
	Bernoulli Naive Bayes (Rahman & Hossen, 2019)	0.8750
	Radial Basis Function (Maulana et al., 2020)	0.8305
IMDb	**Proposed Psychology-Informed Approach**	**0.9053**
	ALBERT (Ding et al., 2021)	0.8905
	LSTM (Qaisar, 2020)	0.8990
	Gate Recurrent Unit (Ding et al., 2021)	0.8631
	Logistic Regression (Qaisar, 2020)	0.8914
	Max Entropy Random Forest (Das & Chakraborty, 2018)	0.8991
Twitter	**Proposed Psychology-Informed Approach**	**0.8553**
	Gaussian Naive Bayes (Celebi, 2023)	0.7722
	Multinomial Naïve Bayes (Deshpande & Rao, 2017)	0.8300
SDCNL	**Proposed Psychology-Informed Approach**	**0.7177**
	Support Vector Machine (Gupta et al., 2023)	0.7000
	Logistic Regression (Gupta et al., 2023)	0.7100
	BERT-Dense (Haque et al., 2021)	0.7050
	BERT-Bidirecttional LSTM (Haque et al., 2021)	0.7150
	Mental FLAN (Large Language Model) (Xu et al., 2023)	0.6770

Our findings suggest we can achieve effective results without training additional complex models. We can reduce costs and training time by leveraging simple methods or existing classification models already integrated within the system. This approach enhances explainability, providing clearer insights into classifications without adding more hidden layers for complexity.

CONCLUSION

In conclusion, the integration of personality and emotion-aware features into natural language understanding models, particularly for sentiment analysis and depression detection, has been validated through various experiments and statistical hypothesis testing, demonstrating their effectiveness in enhancing model performance. This approach improves model accuracy by up to 3.4% in sentiment analysis (IMDb dataset) and depression detection (SDCNL dataset). However, it also enhances model explainability, making it more transparent and interpretable compared to a complex model.

While the model demonstrated improved accuracy without relying strongly on complexity, it includes some features that may not contribute significantly. Therefore, we suggest exploring techniques such as principal component analysis (PCA) or other feature selection methods in future work to eliminate unimportant features and further optimize model performance.

This research has established a foundation for improving explainability and could be expanded to examine additional psychological factors or explore different natural language understanding (NLU) tasks. The future work in this area holds exciting possibilities. For example, future studies could use this approach to assess public intentions regarding vaccination or gauge people's perceptions of policy-making. All these applications aim to enhance public welfare while avoiding the complexity of advanced models. Subsequently, the next phase of work could involve incorporating psychology-informed model features into embeddings without relying solely on recognition results. This approach aims to explore alternative methods of leveraging the model's insights.

ACKNOWLEDGEMENT

The authors would like to express their sincere appreciation for the support provided by the Universiti Tunku Abdul Rahman Research Fund (IPSR/RMC/UTARRF/2021-C1/K03) and Greatech Integration (M) Sdn. Bhd. under project number 8084/0006.

REFERENCES

Adoma, A. F., Henry, N. M., & Chen, W. (2020). Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In *2020 17th International Computer Conference on Wavelet Active Media*

Technology and Information Processing (ICCWAMTIP) (pp. 117-121). IEEE. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379

Aizawa, A. (2003). An information-theoretic perspective of TF–IDF measures. *Information Processing & Management*, *39*(1), 45–65. https://doi.org/10.1016/S0306-4573(02)00021-3

Alfarizi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional text classification using TF-IDF (term frequency-inverse document frequency) and LSTM (long short-term memory). *JUITA: Jurnal Informatika*, *10*(2), 225-232. https://doi.org/10.30595/juita.v10i2.13262

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356-366. https://doi.org/10.1037/0022-3514.86.2.356

Celebi, M. E. (2023). *Nlp for depression prediction*. Kaggle. https://www.kaggle.com/code/merencelebi/nlp-for-depression-prediction

Cervone, D. & Pervin, L. A. (2022). *Personality: Theory and research*. John Wiley & Sons. https://doi.org/10.48550/arXiv.1806.06407

Das, B., & Chakraborty, S. (2018). *An improved text sentiment classification model using TF-IDF and next word negation*. arXiv:1806.06407. https://doi.org/10.48550/arXiv.1806.06407

Deimann, R., Preidt, T., Roy, S. & Stanicki, J. (2023). *Is personality prediction possible based on Reddit comments?* arXiv:2408.16089. https://arxiv.org/abs/2408.16089

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G. & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4040–4054). Association for Computational Linguistics.

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). *GoEmotions: A dataset of fine-grained emotions*. arXiv:2005.00547. https://doi.org/10.48550/arXiv.2005.00547

Deshpande, M., & Rao, V. (2017). Depression detection using emotion artificial intelligence. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 858–862). IEEE. https://doi.org/10.1109/ISS1.2017.8389299

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics.

Ding, Z., Qi, Y., & Lin, D. (2021). Albert-based sentiment analysis of movie review. In *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)* (pp. 1243-1246). IEEE. https://doi.org/10.1109/AEMCSE51986.2021.00254

Durairaj, A. K., & Chinnalagu, A. (2021). Transformer based contextual model for sentiment analysis of customer reviews: A fine-tuned BERT. *International Journal of Advanced Computer Science and Applications*, *12*(11), 474-480. https://doi.org/10.14569/IJACSA.2021.0121153

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Junior & Adult)*. Hodder and Stoughton Educational. https://doi.org/10.1111/j.2044-8341.1977.tb02414.x

Figuerêdo, J. S. L., Maia, A. L. L. M., & Calumby, R. T. (2022). Early depression detection in social media based on deep learning and underlying emotions. *Online Social Networks and Media*, *31*, Article 100225. https://doi.org/10.1016/j.osnem.2022.100225

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26-34. https://doi.org/10.1037/0003-066X.48.1.26

Gupta, S., Sinha, A. M., Prodhan, D., Ghosh, N., & Modak, S. (2023). Detecting depression and suicidal ideation from texts using machine learning & deep learning techniques. *International Journal of Computer Sciences and Engineering* , *11*(1), 29–35.

Haque, A., Reddi, V., & Giallanza, T. (2021). Deep learning for suicide and depression identification with unsupervised label correction. In I. Farkaš, P. Masulli, S. Otte & S. Wermter (Eds.) *Artificial Neural Networks and Machine Learning – ICANN 2021* (pp. 436–447). Springer. https://doi.org/10.1007/978-3-030-86383-8_35

Hartmann, J. (2022). *Emotion-english-distilroberta-base*. Huggingface. https://huggingface.co/j-hartmann/emotion-english-distilroberta-base

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). *ALBERT: A lite BERT for self-supervised learning of language representations*. arXiv:1909.11942. https://doi.org/10.48550/arXiv.1909.11942

Lin, J., Mao, W., & Zeng, D. D. (2017). Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems*, *132*, 204–214. https://doi.org/10.1016/J.KNOSYS.2017.06.031

Liu, C., Osama, M. & de Andrade, A. (2019). *DENS: A dataset for multi-class emotion analysis*. arXiv:1910.11769. https://doi.org/10.48550/arXiv.1910.11769

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv:1907.11692. https://doi.org/10.48550/arXiv.1907.11692

Lowe, S. (2023). *Roberta base go emotion*. Dataloop https://dataloop.ai/library/model/samlowe_roberta-base-go_emotions/

Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 142-150). Association for Computational Linguistics

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

Maulana, R., Rahayuningsih, P. A., Irmayani, W., Saputra, D., & Jayanti, W. E. (2020). Improved accuracy of sentiment analysis movie review using support vector machine based information gain. *Journal of Physics: Conference Series*, *1641*(1), Article 012060. https://doi.org/10.1088/1742-6596/1641/1/012060

Mitchell, J. (2017). *Myers-briggs personality type dataset*. Kaggle. https://www.kaggle.com/datasets/datasnaek/mbti-type

Moreno, J. D., Martínez-Huertas, J., Olmos, R., Jorge-Botana, G., & Botella, J. (2021). Can personality traits be measured analyzing written language? A meta-analytic study on computational methods. *Personality and Individual Differences*, *177*, Article 110818. https://doi.org/10.1016/J.PAID.2021.110818

Pang, B., & Lee, L. (2004). *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*. arXiv:cs/0409058. https://doi.org/10.48550/arXiv.cs/040905

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Qaisar, S. M. (2020). Sentiment analysis of IMDb movie reviews using long short-term memory. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICCIS49240.2020.9257657

Rahman, A. & Hossen, Md. S. (2019). Sentiment analysis on movie review data using machine learning approach. *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 1–4. https://doi.org/10.1109/ICBSLP47725.2019.201470

Rajat, R., Jaroli, P., Kumar, N., & Kaushal, R. K. (2021). A sentiment analysis of amazon review data using machine learning model. In *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)* (pp. 1-6). IEEE. https://doi.org/10.1109/CITISIA53721.2021.9719909

Raw, N. (2021). *Bert base uncased emotion*. Huggingface. https://huggingface.co/nateraw/bert-base-uncased-emotion

Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management*, *58*(3), Article 102532. https://doi.org/10.1016/j.ipm.2021.102532

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), 310-328.

Shaukat, Z., Zulfiqar, A. A., Xiao, C., Azeem, M., & Mahmood, T. (2020). Sentiment analysis on IMDB using lexicon and neural networks. *SN Applied Sciences*, *2*(2), Article 148. https://doi.org/10.1007/s42452-019-1926-x

Shinde, V. (2022). *Twitter dataset*. Kaggle. https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media

Smith, M. Q. R. P., & Ruxton, G. D. (2020). Effective use of the McNemar test. *Behavioral Ecology and Sociobiology*, *74*(11), Article 133. https://doi.org/10.1007/s00265-020-02916-y

Sönmezöz, K., Uğur, Ö., & Diri, B. (2020). MBTI personality prediction with machine learning. In *2020 28th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. https://doi.org/10.1109/SIU49456.2020.9302239

Sun, G., Guo, A., Ma, J., & Wei, J. (2019). Personal trait analysis using word2vec based on user-generated text. In *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 1131-1137). IEEE. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00213

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, *7*, 44883–44893. https://doi.org/10.1109/ACCESS.2019.2909180

Tan, J. J. (2024a). *IMDb sentiment dataset splitting*. Huggingface. https://huggingface.co/datasets/jingjietan/imdb-sentiment

Tan, J. J. (2024b). *MBTI-personalitycafe-MBTI personality dataset splitting*. Huggingface. https://huggingface.co/datasets/jingjietan/kaggle-mbti

Tan, J. J. (2024c). *Polarity sentiment dataset splitting*. Huggingface. https://huggingface.co/datasets/jingjietan/polarity-sentiment

Tan, J. J. (2024d). *SDCNL suicide dataset splitting*. Huggingface. https://huggingface.co/datasets/jingjietan/sdcnl-suicide

Tan, J. J. (2024e). *Twitter depression dataset splitting*. Huggingface. https://huggingface.co/datasets/jingjietan/twitter-depression

Tan, J. J., Kwan, B. H., Ng, D. W. K., & Hum, Y. C. (2023, December 14-15). *Enhancing performance in sentiment classification of textual messages through personality recognition integration*. [Paper presentation]. 11th International Symposium on Applied Engineering and Sciences, Selangor, Malaysia.

Xu, X., Yao, B., Dong, Y., Gabriel, S., Yu, H., Hendler, J., Ghassemi, M., Dey, A. K., & Wang, D. (2023). Mental-llm: Leveraging large language models for mental health prediction via online text data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 8(1), 1-32*. https://doi.org/10.1145/3643540

Yang, W., Yuan, T., & Wang, L. (2020). Micro-blog sentiment classification method based on the personality and bagging algorithm. *Future Internet*, *12*(4), 75. https://doi.org/10.3390/fi12040075

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Yates, A., Cohan, A., & Goharian, N. (2017). Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2968–2978). Association for Computational Linguistics