**STUDENT NAME: BOBOYE DAMILARE ADEJUWON**

**STUDENT NUMBER: 202346817**

**Exploring Natural Language Processing Techniques for Depression Detection: From Logistic Regression and SVM to BiLSTM and Transformers**

**Abstract**

In recent times, individuals take to social media to express their feelings and emotions which might indicate symptoms of depression, albeit difficult to diagnose. Research employing user-generated social media posts has been carried out to detect symptomatic signs of depression. In this paper, machine learning models such as Logistic Regression (LR), Support Vector Machine (SVM), Bidirectional Long-Short-Term Memory (BiLSTM) and attention-based Transformers have been applied to classify posts suggesting depressive language. Extensive preprocessing steps, including tokenization, lemmatization, n-gram modeling, and vectorization using TF-IDF, CountVectorizer, and Word2Vec embedding, were applied to generate rich contextual meaning and capture semantic relationships. The attention-based Transformer model achieved a ROC-AUC score of 0.99 and an F1-score of 0.94, emerging as the most effective model after fine-tuning. The results supplement the prospect of NLP in detecting depression early, specifically through the use of deep learning architectures capable of capturing complex linguistic patterns in informal discourse.

**Introduction**

* 1. **Definition**

Globally, depression is estimated to affect 280 million people (Chen et al., 2023). Early detection enables timely intervention and the provision of necessary mental health services, significantly improving outcomes and potentially preventing severe consequences, such as suicide.

A subspeciality of Artificial Intelligence, Natural Language Processing (NLP), provides an analytical framework for analysing and interpreting human language. Applied to textual data, NLP can unearth linguistic patterns and identify language markers indicative of depressive symptoms, making automated early detection systems feasible.

The proliferation of microblogging platforms has enabled users to openly express thoughts, feelings, and emotional struggles through written narratives. Therefore, this study contributes to the NLP domain by identifying and classifyingdepression-indicative language in user-generated text, particularly within non-clinical, informal online discourse.

**1.2 Scope**

This report aims to improve the detection of depression by uniquely identifying specific words in written texts indicative of depressive symptoms. As submitted by Triantafyllopoulos et al. (2023), these expressions, often informal or coded in non-clinical terms, present a unique challenge: **Depression-related sentiments are resistant to direct clinical quantification due to their nuanced, context-dependent, and emotionally ambiguous nature**.As such,these linguistic markers can provide deeper insight into individuals’ psychological states.

**1.3 Importance**

According to an article by the Financial Times (2024), about $1 trillion is lost each year in the global economy due to lost working days as a result of depression and anxiety disorders. Hence, applying NLP techniques to social media data to mitigate these impacts will result in the timely detection of depressive symptoms.

Recent studies highlight the impact of combining NLP techniques with linguistic features that convey emotional states. However, due to the nuanced ways individuals express emotions, especially in informal platforms like social media, it becomes challenging to predict mental health states, requiring further research.

**Review of Related Studies**

In this section, a background review of past studies is conducted by searching peer-reviewed databases such as Google Scholar, PubMed, IEEE Xplore, Science Direct, and others for related studies. Also, arXiv, an open-access repository, was canvassed. Keywords such as ‘Natural Language processing for Depression Detection’ were used to filter the databases, and relevant studies were selected.

Chen et al. (2023) applied hybrid deep learning approach to reveal individuals with depression by analyzing their Reddit posts. Merging Sentence BERT (SBERT) and Convolutional Neural Network (CNN) for semantic representation and temporal patterns, their model achieved an F1 score of 0.86, outperforming previous models in the literature. The model’s complexity effectively captures semantic representations and temporal behavioral patterns, enhancing detection capabilities. However, this complexity could pose a challenge in real-time deployment due to computational resource issues.

Inamdar et al. (2023) employed various NLP techniques, like Bag of Words (BoW), Word2Vec, ELMo (Embeddings from Language Models), and BERT (Bidirectional Encoder Representations from Transformers), to develop traditional machine learning models to detect mental health stress using Reddit posts. The synergy of ELMo embeddings with Support Vector Machine (SVM) produced the best result - an F1 score of 0.76. Although authors reported small dataset as a limitation of the study, the study applied state-of-the-art NLP techniques to preprocess and generate sentence data before being fed into traditional machine learning models, achieving high performance.

​Triantafyllopoulos et al. (2023) developed a deep learning architecture for identifying depression within social media content, integrating emotional and socio-normative features with BERT embeddings and an attentive bidirectional GRU network through a late fusion approach. Their model underwent evaluation on two datasets: the Pirina corpus, focusing on post-level classification, and the Reddit Self-reported Depression Diagnosis (RSDD) corpus, focusing on user-level classification. Including affective and social norm features led to the model yielding absolute F1 score improvements of 2.65% and 6.73% on the former and latter datasets, respectively. While these enhancements improved the model's performance, potential challenges in generalizing the model across diverse populations may arise.

Bucur et al. (2023) developed a temporally enriched multimodal transformer model using Twitter and Reddit multimodal datasets. By employing pretrained models to draw out image and text embeddings, using CLIP-based image and EmoBERTa text embeddings with time2vec positional encodings, their proposed model achieved F1 scores of 0.931 and 0.902 on both datasets, surpassing existing methods. The capacity of the model to process multimodal data in a time-sensitive context makes it robust, allowing for user behavior monitoring over time. However, the model’s complexity may hinder scalability since it requires multimodal data, which may not always be available.

Lastly, Teferra et al. (2024) surveyed natural language processing techniques, including sentiment analysis, linguistic biomarkers for depression detection, and concluded that traditional machine learning models and deep learning models, such as language models (i.e., Transformers), offer excellent avenues for developing depression screening. While their review highlights the strengths and limitations of various NLP techniques, they identified and addressed ethical issues and cross-cultural perspectives.

**2.1 SMART Objectives**

SMART objectives for this Natural Language Processing task are highlighted in detail as follows:

**Specific Objective**

The specific objective of this project is to build traditional machine learning and deep learning methods to classify whether an individual is depressed or not depressed, based on textual data obtained from Reddit.

**Measurable Objective**

The models will be evaluated using at least two standard classification metrics, including F1-score, and ROC-AUC (Receiver Operating Characteristic – Area Under Curve). Based on the literature, the goal is to achieve an F1-score of 90% for the deep learning models, with a comparable performance from traditional models. The aim is to surpass the existing benchmarks in terms of F1-score, as achieved in the study by Inamdar et al. (2023).

**Achievable Objective**

This objective is based primarily on existing research in mental health prediction in the Natural Language Processing (NLP) domain. The NLP techniques utilized in this study have been successfully applied in prior studies and on similar datasets for classifying depression.

**Relevant Objective**

This project attempts to address the growing need for digital mental health monitoring, specifically for depression, on social media outlets. As individuals take to social media their feelings and emotions, the need to automatically detect symptoms of depression from texts can significantly contribute to early intervention efforts.

**Time-Bound Objective**

This project will be completed over two weeks, with the key milestones as follows:

* Data collection and preprocessing by the 13th of April
* Model development and evaluation by the 23rd of April
* Final report submission by the 29th of April

This timeline ensures the project is completed with adequate time for refinement and documentation of results.

**Methodology**

This section includes information on the dataset, exploratory data analysis, natural language processing techniques, traditional machine learning models, and deep learning models. Also includes the implementation and refinement of various models.

**3.1 Dataset**

The dataset for this project has been obtained from a popular data science platform called Kaggle. It is publicly available. It contains textual corpus web-scraped from Reddit, a popular microblogging site, and is suitable for depression detection analysis using Natural Language Processing.

The dataset structure is tabular, containing two columns, namely:

* **clean\_text:** Preprocessed Reddit posts containing posts (texts) generated by users.
* **is\_depression:** a binary label that indicates whether a post exhibits signs of depression, denoted as 1, and not depressed, denoted as 0.

The data size is 15462 rows and is entirely textual. The shape is (7731, 2), where each row represents an individual Reddit post with an assigned label. The dataset is appropriate for the task at hand – depression detection through textual analysis. It provides a realistic and challenging binary classification problem.

As shown in Figure 1, the dataset class is balanced with 3900 rows belonging to class 0, and 3831 belonging to class 1.

A green and blue rectangular shapes

AI-generated content may be incorrect.**Figure 1: Mental Health Status Labels**

Figure 2 shows the Natural Language Task workflow which was inspired by Inamdar et al (2023).

A screenshot of a computer

AI-generated content may be incorrect.

**Figure 2: Workflow Diagram**

**3.2 Exploratory Data Analysis**

In this section, the exploratory data analysis techniques applied to preprocess the data are explained thoroughly.

**Data Cleaning:** Data was examined for missing values and duplicated values. Duplicated values in the label were dropped.

**Data Preprocessing:** This process involved tokenization, lemmatization, removing special characters, punctuation marks, hyperlinks like HTML tags, gibberish texts, texts that are not in the English language, and converting all sentences to lowercase, which may impact classification and text analysis.

Furthermore, Natural Language Processing techniques for data preprocessing are described as follows:

* Text preprocessing: Words in the corpus without significant information were removed using the stopwords library from NLTK. The re package from Python was used to remove punctuation marks, hashtags, and other special characters. Additionally, texts in the corpus were converted to lowercase to improve classification and text analysis.
* Text Tokenization: Texts were tokenized, converting raw string inputs into vectors suitable for embedding layers. Furthermore, Term Frequency-Inverse Document Frequency (TF-IDF) and CountVectorizer were applied to convert texts into numerical matrices and quantify the importance of specific words.
* N-gram construction: Enhancing the semantic representation of words involved creating n-gram modeling. Specifically, bigram models (two-word combinations) and trigram models (three-word combinations). These models identify frequently co-occurring word sequences and merge them into single tokens, capable of capturing meaningful multi-word expressions such as ‘*mental illness’* and ‘*intrusive thought’.*
* Lemmatization: Lemmatization was applied to extract the root meaning of words in our corpus. This process improves the reduction of inflectional forms and derivationally related words to a common base. For example, words such as ‘running’ become ‘run’. Also, we employed the spaCy language model on a list of tokenized sentences. It provided part-of-speech tagging and grammatical structure analysis to ensure context-aware lemmatization since the same word may serve different grammatical roles.

Following thorough data preprocessing, a word cloud representing the frequency of words in the corpus was generated as shown in Figure 3.

**A collage of words

AI-generated content may be incorrect.**

**Figure 3: WordCloud of Corpus**

Also, the text length distribution was visualized with the highest sentence length reaching as much as 4000 words, as shown in Figure 4. A graph of a number of words

AI-generated content may be incorrect.

**Figure 4: Distribution of Text Length**

**Baseline Model Selection**

Considering the literature review, the baseline model for this study was selected as a logistic regression classifier. It is a widely adopted linear model for tasks involving two-class classification, with robust performance across a variety of NLP applications, including text classification, spam detection, and mental health detection.

**3.3 Comparison of Traditional Machine Learning Models**

Five traditional machine learning models are compared and evaluated based on their strengths and weaknesses and suitability for the depression detection task using natural language inputs. These models were chosen due to their prior use in mental health classification research and the domain of Natural Language Processing for other classification tasks.

**Logistic Regression (LR):** Logistic regression is frequently used for tasks requiring classification where the possible number of outcomes are two. It employs a sigmoid function to calculate the probability that a discrete variable will be assigned to a particular category (Aliman et al., 2022) and is suited for mental health classification tasks due to its interpretability and simplicity. Although its performance can be limited with complex datasets because it relies on a linear decision boundary.

**Random Forest (RF):** This ensemble model aggregates several decision trees trained on different parts of the data (Breiman, 2001). They are employed for classification and regression tasks and perform well with heterogeneous data and are robust to overfitting. However, due to their computational complexity and lack of interpretability compared to linear models, they can be opaque, making them a disadvantage.

**Multinomial Naive Bayes (MNB)**  
Rooted in Bayes’ Theorem with a strong assumption of feature independence, particularly when input features are term frequencies or TF-IDF values (McCallum & Nigam, 1998). It assumes that features (words) follow a multinomial distribution and occur independently. While this feature independence assumption is rarely true in language, MNB often performs well in practice and is especially efficient on large datasets.

**Decision Tree (DT):** A tree-like model that splits features into regions based on decision support – information gain or entropy measure (Quinlan, 1986). While they are highly interpretable due to their visualization component and adaptable to various data types, they tend to overfit, especially on sparse, high-dimensional NLP datasets, and require pruning and regularization.

**Support Vector Machine (SVM):** Support vector machine is effective in a high-dimensional data space, such as natural language processing. They are utilized for text categorization applications and are robust to outliers (Joachims, 2002). With the use of kernel functions like radial basis function (RBF), they are able to capture non-linear boundaries. However, their scalability is limited, especially with large datasets or dense word embeddings.

Based on comparative analysis, logistic regression and support vector machine were chosen as the primary traditional machine learning models for this study. Logistic regression provides a strong baseline, and its wide application in binary classification tasks and low computational cost make it suitable for depression detection. Support vector machine was selected due to its performance in text classification involving high-dimensional inputs such as word embeddings. Its ability to capture non-linear decision boundaries via kernel functions makes it particularly powerful for complex linguistic features associated with depression detection.

**3.4 Comparison of Deep Learning Methods**

Recently, deep learning approaches have become pivotal in natural language processing due to their capacity to model relationships between syntactic patterns and context using vectors. In the context of this report, we compare five deep learning methods.

**Long Short-Term Memory (LSTM):** A unique type of Recurrent Neural Network (RNN) that has memory cell with gating mechanisms - an input, forget, and output gate, which it uses to control information flow as it propagates through different time steps (Berrajaa, 2022). Because they can learn long-range dependencies, they defeat the vanishing gradient problem in RNN, making them suitable for sentiment analysis and depression detection. However, they are computationally expensive, have memory issues, and struggle with very long sentences.

**Gated Recurrent Unit (GRU):** A simplified LSTM architecture that combines the input and forget gate into an update gate. The network controls information flow by selectively remembering or forgetting relevant sequences, reducing complexity. While they are suitable for our textual analysis task, they struggle with long-term dependencies compared to LSTM.

**Bidirectional LSTM (BiLSTM):** In comparison to conventional LSTMs that process sequential information in only a forward direction, BiLSTMs are an extension in that they can process sequences in forward and reverse directions, allowing for richer contextual understanding. This dual representation is particularly advantageous in depression detection, where subtle cues may depend on both preceding and following words. While more computationally expensive, requiring longer training time than standard LSTMs or GRUs, BiLSTMs often outperform them in text classification tasks (Asrawi et al., 2023).

**Attention-based Transformers:** Developed by Google researchers in 2017, attention-based transformers have a complex architecture; however, they defeat the memory issue (vanishing gradient) experienced by recurrent neural networks. Introduced by Vaswani et al. (2017), they adopt self-attention mechanisms to model dependencies among words, with respect to their relative distance in the text. As such, this allows for highly parallelizable training and effective long-range context modeling, as required, where the longest sentence in our corpus is more than 4000 words. While highly suitable for modeling textual data, they are computationally expensive.

**Bidirectional Encoder Representations from Transformers (BERT):** BERT is a self-supervised learning algorithm that uses only an encoder transformer architecture. It has been pretrained on a large corpus and thus adequately captures semantic and context-based meanings from sentences. While it is suitable for the depression detection task at hand, it requires careful fine-tuning.

In light of the comparative analysis, attention-based transformers and bidirectional LSTM have been chosen for this study. BiLSTM’s ability to capture semantic meaning and relationship between words in both directions makes it suitable for depression detection, where subtle cues may depend on both preceding and following words. Similarly, attention-based transformers capture the rich semantic contextual relationship between words and long-range dependencies without the sequential limitations of recurrent neural networks. Their self-attention mechanisms are particularly effective in highlighting emotionally charged terms or patterns indicative of depressive states (Nadeem et al., 2022).

**3.5 Implementation and Refinement**

This section describes the libraries and procedures used in building, training, and testing various machine learning pipelines.

* **NLTK, Gensim, spaCy:** These are Python libraries for Natural Language Processing (NLP). They were used to preprocess the raw text data through tokenization, stopword removal, lemmatization, and part-of-speech tagging. Additionally, Gensim was employed to train Word2Vec embeddings, which provided pretrained vector representations of words. The embeddings were fed as input features for BiLSTM, enhancing the model’s understanding of semantic relationships between words.
* **Scikit-learn:** A robust machine learning library offering tools for model development, evaluation, and preprocessing (Ngunyen et al., 2019). CountVectorizer and TF-IDF were used to convert text into numerical feature matrices for Logistic Regression and Support Vector Machines (SVM) which were trained on 70% of the dataset and evaluated on the remaining 30%. GridSearchCV was applied for hyperparameter tuning to determine optimal parameters to enhance model performance and generalizability.
* **TensorFlow:** This is a publicly available library that can perform high computation and high-scale machine learning tasks (Ngunyen et al., 2019), developed for deep learning models, including a Bidirectional LSTM (BiLSTM) and an Attention-based Transformer model. Both were trained on 70% and tested on the remaining 30% of the data. To fine-tune the performance and prevent overfitting, several strategies were adopted:
  1. **Hyperparameter tuning:** BiLSTM and attention-based Transformer architecture units were adjusted. Also, learning rate and batch size were also modified iteratively to improve the model’s performance.
  2. **Regularization:** L2 regularization was applied to the fully connected layers to minimize large weights and reduce overfitting.
  3. **Early stopping:** A technique where training is stopped when validation performance no longer improves, thereby preventing overfitting, reducing training time.

The architecture of the BiLSTM Model is described in Table 1**.**

**Table 1: BiLSTM Architecture.**

|  |  |
| --- | --- |
| **Layer Type** | **Details** |
| Input layer | Sequence of tokenized words |
| Embedding Layer | Pretrained Word2Vec (300-dim), non-trainable, input\_length = 1000 |
| Bidirectional LSTM – first layer | 64 units |
| Bidirectional LSTM – second layer | 64 units |
| Dense Output layer | 1-unit, sigmoid activation function (binary classification) |
| Optimizer | Adam (default learning rate) |
| Vocabulary size | 10000 |
| Loss function | Binary Crossentropy |
| Training parameters | Batch size = 32, Epoch = 10. |

The architecture of the attention-based Transformer is described in Table 2.

**Table 2: Attention-based Transformer architecture.**

|  |  |  |
| --- | --- | --- |
| **Layer** | **Output shape** | **Description** |
| Input | None, maxlen | Accepts sequences of token indices with max length (1000) |
| Token and Position embedding - Custom Embedding layer | None, maxlen, embed\_dim | Merges token and positional embeddings to represent input sequence |
| Transformer Block – Custom Transformer Layer | None, maxlen, embed\_dim | Utilizes multi-head self-attention and feed-forward network |
| Global Average Pooling – 1D | None, embed\_dim | Aggregates sequence information by averaging over the time dimension |
| Dropout | None, embed\_dim | Applies dropout rate of 0.1 to prevent overfitting |
| Dense | None, 20 | Fully connected layer with ReLU activation function that introduces non-linearity to the network |
| Dropout | None, 20 | Applies a dropout rate of 0.1. |
| Output – Dense Layer | None, 1 | Final output layer with sigmoid activation function for binary classification |
|  |  |  |
| **Training parameters** | **Description** | - |
| Loss function | Binary Crossentropy | - |
| Optimizer | Adam (Default learning rate) | - |
| Batch size | 32 | - |
| Epoch | 10 | - |
| Metrics | Accuracy | - |

**Evaluation**

This section discusses the evaluation of various models using different metrics and visualizations such as F1-score, ROC-AUC plots, tables, and confusion matrices. Table 3 below presents the confusion matrix and ROC curve for LR and SVM.

**Table 3: Evaluation Metrics for LR and SVM.**

|  |  |
| --- | --- |
| A screenshot of a graph  AI-generated content may be incorrect. | A graph of a logistic regression  AI-generated content may be incorrect. |
| A screenshot of a graph  AI-generated content may be incorrect. | A graph of a function  AI-generated content may be incorrect. |

As shown in Table 3, SVM achieved a slightly higher accuracy (0.94) than Logistic Regression (LR), though both models recorded identical ROC-AUC scores of 0.97, indicating a strong ability to differentiate class labels. F1-scores were 0.93 (non-depressed) and 0.94 (depressed), surpassing the 80% benchmark reported by Inamdar et al. (2023). After hyperparameter tuning via GridSearchCV, LR outperformed SVM on both accuracy and F1-score, as presented in Table 4.

**Table 4:** **Evaluation Metrics for Finetuned LR and SVM.**

|  |  |
| --- | --- |
|  |  |

The deep learning models, BiLSTM and Attention-based Transformers, were evaluated, and their training vs validation loss is presented in Table 5.

**Table 5: BiLSTM and Attention-based Transformers Model Accuracy vs Loss over Epochs (Training).**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

As shown in Table 5, the BiLSTM model learns well on the training data and will adapt to previously unseen data, however, there are signs of overfitting as the epoch increases, especially when finetuned, the BiLSTM model performs poorly than before, which might indicate further parameter tuning.

The Attention-based Transformer model performs very well on the training data with high accuracy and minimal loss. Moreover, when finetuned, the model performs significantly better than before, achieving stronger generalization with very minimal indications of overfitting.

**Table 6: F1-score and ROC-AUC Evaluation of BiLSTM and Attention-based Transformers**

|  |  |  |
| --- | --- | --- |
| **Model Performance Before Finetuning** | **F1-score** | **ROC-AUC** |
| **BiLSTM** | **0.95** | **0.99** |
| **Attention-based Transformer** | **0.94** | **0.98** |
|  |  |  |
| **Model Performance After Finetuning** | **F1-score** | **ROC-AUC** |
| **BiLSTM** | **0.85** | **0.95** |
| **Attention-based Transformer** | **0.94** | **0.99** |

**Table 7: Confusion Matrix for** **BiLSTM and Attention-based Transformers**

|  |  |
| --- | --- |
|  |  |
|  |  |

As shown in Tables 6 and 7 above, the attention-based transformer model outperforms the BiLSTM model on F1-score (0.94) and ROC-AUC (0.99) when finetuned. The confusion matrix shows that the finetuned attention-based transformer model excels at ranking positive cases higher than negative ones.

**Conclusion**

This study has analysed textual corpus obtained from Reddit for depression detection using traditional and deep learning models. It employed TF-IDF and CountVectorizer for preprocessing traditional machine learning models like LR and SVM, with the logistic regression model outperforming the latter when finetuned with an accuracy of 0.93. Furthermore, word2vec embeddings were used for the BiLSTM model, and Keras tokenizer for the Attention-based Transformer model, with the attention-based transformer model slightly outperforming the BiLSTM with an F1-score of 0.94 and ROC-AUC of 0.99 when finetuned using L2 regularizer and early stopping with patience of 3 for validation loss and a batch size of 64.

The reference style followed in this paper is the Harvard citation style that uses author-date references compared to Vancouver, which is often used in medical research, Chicago style which uses two systems – notes and bibliography, and the preferred APA (American Psychological Association) in social sciences, and finally the MLA (Modern Language Association) often used in the humanities. Future research will include more textual corpora from other social media platforms. Limitations noted are small and diverse datasets.

**References**

Aliman, G., Nivera, T.F.S., Olazo, J.C.A., Ramos, D.J.P., Sanchez, C.D.B., Amado, T.M. and Valenzuela, I.C., 2022. Sentiment analysis using logistic regression. *Journal of Computational Innovations and Engineering Applications*, *7*(1), pp.35-40.

Asrawi, H., Utami, E. and Yaqin, A., 2023. The LSTM and Bidirectional GRU Comparison for Text Classification. *Sinkron: jurnal dan penelitian teknik informatika*, *7*(4), pp.2264-2274.

BERRAJAA, A., 2022. Natural language processing for the analysis sentiment using a LSTM model. *International Journal of Advanced Computer Science and Applications*, *13*(5).

Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

Bucur, A.M., Cosma, A., Rosso, P. and Dinu, L.P., 2023, March. It’s just a matter of time: Detecting depression with time-enriched multimodal transformers. In *European conference on information retrieval* (pp. 200-215). Cham: Springer Nature Switzerland.

Chen, Z., Yang, R., Fu, S., Zong, N., Liu, H. and Huang, M., 2023, June. Detecting Reddit users with depression using a hybrid neural network SBERT-CNN. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)* (pp. 193-199). IEEE.

Financial Times. (2024). The Big Question? What is driving the Global Mental Health Crisis, *Financial Times*, 10 October. Available at: <https://www.ft.com/content/4b75300a-12ca-42b7-a07a-4519435ea4b8> (Accessed: 22 April 2025).

Haque, A., Reddi, V. and Giallanza, T., 2021. Deep learning for suicide and depression identification with unsupervised label correction. In *Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part V 30* (pp. 436-447). Springer International Publishing.

Inamdar, S., Chapekar, R., Gite, S. and Pradhan, B., 2023. Machine learning driven mental stress detection on reddit posts using natural language processing. *Human-Centric Intelligent Systems*, *3*(2), pp.80-91.

Joachims, T., 2002. *Learning to classify text using support vector machines* (Vol. 668). Springer Science & Business Media.

McCallum, A. and Nigam, K., 1998, July. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).

Nadeem, A., Naveed, M., Islam Satti, M., Afzal, H., Ahmad, T. and Kim, K.I., 2022. Depression detection based on hybrid deep learning SSCL framework using self-attention mechanism: An application to social networking data. *Sensors*, *22*(24), p.9775.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., Malík, P. and Hluchý, L., 2019. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, *52*, pp.77-124.

Quinlan, J.R., 1986. Induction of decision trees. *Machine learning*, *1*, pp.81-106.

Teferra, B.G., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan, S. and Bhat, V., 2024. Screening for Depression Using Natural Language Processing: Literature Review. *Interactive Journal of Medical Research*, *13*(1), p.e55067.

Triantafyllopoulos, I., Paraskevopoulos, G. and Potamianos, A., 2023. Depression detection in social media posts using affective and social norm features. *arXiv preprint arXiv:2303.14279*.