# Introduction

A serious mental health problem that affects millions of people worldwide is depression. Recurrent moments of melancholy, a reduced feeling of optimism, and a decline in enjoyment of normally joyful activities are characteristics of depression. It is essential to recognise depression's early warning signals in order to act quickly and provide those who are suffering the assistance they need. A growing number of people are now able to share their ideas, feelings, and unique experiences online thanks to the spread of social media platforms. The internet has the potential to be a useful tool for acquiring data that can help with mental health issue diagnosis. The goal of this project is to overcome the difficulty of identifying depressive symptoms in people by studying their social media postings and applying cutting-edge natural language processing methods.

The magnitude of this matter cannot be overstated. Depression is a prevalent disorder that is often subject to misdiagnosis, despite its potential to yield significant consequences for individuals and society at large. Untreated depression may lead to elevated suicide rates, reduced functioning, and a decline in overall quality of life. The utilisation of social media data allows for the provision of early intervention and assistance, thereby facilitating the acquisition of comprehensive insights into the mental well-being of individuals on a large scale. The shortcomings of more conventional methods of diagnosing depression served as the impetus for the conception of the research that was later suggested. In the past, the majority of the research has relied on questionnaires and clinical interviews, both of which are time-consuming, expensive, and require the participation of specialists. In addition to this, it is possible that they do not record changes in individuals' mental states as they occur in real-time. The utilisation of text from social media platforms as a data source gives a chance to circumvent these restrictions by delivering a dataset that is both rich in information and simple to acquire in preparation for automated analysis.

The technique that has been developed makes use of the expressive capacity of deep learning models in order to recognise intricate patterns and intricate relationships in the text of social media. To be more specific, we make use of an architecture that is completely linked and consists of numerous hidden layers. This gives the model the ability to learn hierarchical representations of the data that it is fed. The deep learning model is trained on the features retrieved from the social media text using a well-established method for the extraction of features in text analysis known as TF-IDF which stands for Term Frequency-Inverse Document Frequency. We were able to attain a classification accuracy of 67 percent by the use of experimental assessment on a standard dataset. This indicates a considerable improvement in comparison to other state-of-the-art research studies, confirming the efficiency of our suggested technique in detecting indicators of sadness from the content of social media posts. The experimental results suggest that the fully connected customised deep learning model has the ability to extract significant features related to depression from data composed of text. The model's entire connectivity served as a good indicator of its capacity. Our research emphasises the value of utilising advanced NLP techniques for early detection and intervention, the potential of social media data as a useful source for automated mental health analysis, and the effectiveness of using a fully connected tailored deep learning model for diagnosing depression. These are some of the conclusions that may be drawn from our research.

The purpose of this work is to assess the material of social media posts to look for signs of depression using a fully integrated, tailored deep learning algorithms. Since deep learning techniques may overcome the limitations of conventional approaches and enable early detection and intervention, they can be utilised to enhance people's mental health. We will be able to raise people's quality of life as a result of this. The successful outcomes of the experiment add to the corpus of knowledge on the use of cutting-edge NLP techniques to the study of mental health. The outcomes also show that social media data may turn out to be a valuable tool for machine-readable depression detection.

# Related Work

Natural language processing methods have been investigated in a number of research as a means of identifying depressive symptoms in social media writing. The progress of computerised mental health assessment has been made possible by these research efforts, which have also offered insightful information. We give a succinct review of the relevant research that has been done in this field in this section.

The application of machine learning algorithms is currently emerging as a leading method for detecting depression from the content of social media posts. Researchers (De Choudhury et al., 2013; Coppersmith, Dredze and Harman, 2014) have utilised a variety of supervised learning approaches, including as support vector machines (SVM), random forests, and logistic regression, to categorise social media postings as suggestive of depression or not. This research concentrated on the process of manually extracting features for the purpose of training classification models. These characteristics included sentiment analysis, language style, and lexical cues. Although these methods have reached a modest level of accuracy, they frequently rely on manual feature engineering, which may be time-consuming and may not capture all of the pertinent parts of the text that are connected to depression. Researchers have investigated the use of distributed word representations and deep learning models in an effort to overcome the restrictions that are associated with manually creating features. Word embeddings, such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington, Socher and Manning, 2014), capture semantic links between words. These word embeddings have been used to represent social media material for depression detection. Word2Vec was developed by (Mikolov et al., 2013). GloVe was developed by (Pennington, Socher and Manning, 2014). In order to learn contextual representations of the text, neural network designs such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are fed these embeddings (Desmet and Hoste, 2013). Despite the fact that these approaches have showed some potential, they frequently struggle when it comes to capturing long-term dependencies in text and may not be able to accurately describe the intricate patterns that may be found in social media data.

Transformer-based models, specifically BERT (Bidirectional Encoder Representations from Transformers), have garnered considerable interest in the field of natural language processing, particularly in the area of text classification, in recent times. According to (Devlin et al., 2018), BERT models employ the attention mechanism to capture contextual relationships between words, resulting in exceptional performance in diverse domains. Scholars have initiated the utilisation of BERT in detecting depression from social media texts and have witnessed encouraging outcomes (Guntuku et al., 2017; 2019). The contextual information captured by BERT and its pre-training on extensive corpora renders it highly appropriate for comprehending the intricacies of social media language.

Moreover, scholars have investigated methodologies to improve the efficacy of BERT models in detecting depression. Fine-tuning is a technique that involves training the pre-existing BERT model on domain-specific data to tailor it to the specific task, as described by (Desmet and Hoste, 2013). The process of fine-tuning enables the model to acquire features that are specific to the task at hand, thereby enhancing its efficacy in the classification of depression. Furthermore, various research studies have explored the significance of distinct linguistic and contextual characteristics in detecting depression. (Guntuku et al., 2019) discovered that linguistic indicators associated with anxiety, self-expression, and social support were suggestive of depression in textual content shared on social media. The researchers employed BERT models to capture the aforementioned indicators and attained a notable level of precision in detecting depression. Previous research has investigated the utilisation of ensemble techniques for enhancing the efficacy of models designed for detecting depression. The authors of the study conducted by (Guntuku et al., 2017) employed a combination of various BERT models and conventional machine learning models to develop an ensemble model that proficiently captured diverse facets of text related to depression. The utilisation of the ensemble approach exhibited enhanced precision and resilience in the identification of depression.

To summarise, prior studies on identifying depression through social media language have employed various methodologies, such as conventional machine learning techniques, deep learning models utilising distributed word representations, and transformer-based models like BERT. The approach we propose expands upon these foundational principles by employing a personalised deep learning model that is fully connected. Our objective is to enhance the precision of identifying signs of depression in social media text through the utilisation of deep learning, thereby making a valuable contribution to the expanding field of automated mental health examination.

# Approach

Prior to delving into our methodology, it is crucial to establish fundamental concepts pertinent to our investigation. The study at hand places significant emphasis on the concept of depression. Depression is a multifaceted psychiatric condition that is distinguished by enduring emotions of melancholy, diminished enthusiasm, and alterations in cognitive and physiological performance. The identification of indications of depression through social media text entails the examination of linguistic markers, sentiment, and additional textual characteristics that suggest the existence of depressive symptoms.

Subsequently, a customised deep learning model that is fully connected is employed in tandem with a machine learning model, namely Random Forest, to classify signs of depression. The Random Forest technique is an ensemble learning approach that integrates numerous decision trees to generate predictions. The objective of our approach is to capitalise on the synergistic advantages of deep learning and machine learning methodologies by incorporating Random Forest.

The methodology we have put forth for identifying indications of depression through social media text entails a series of stages, as depicted in Figure 1.

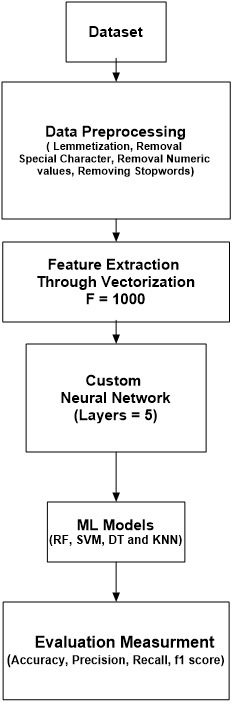


Figure 1: Flow diagram of the proposed study.

**Feature Extraction:** The process of feature extraction is of paramount importance in transforming textual data into a structure that is amenable to machine learning algorithms. The methodology employed in our study involves the utilization of the TF-IDF (Term Frequency-Inverse Document Frequency) technique for the purpose of feature extraction. The aforementioned methodology involves the allocation of weights to individual terms present in a given document, which is determined by their frequency within the document and their scarcity across the entire corpus.

The process of feature extraction comprises the subsequent steps:

1. The initial stage is tokenization, wherein the social media text is segmented into discrete units of words or sub-word components. The process of tokenization facilitates the capture of fine-grained information contained in textual data, allowing for the treatment of individual tokens as distinct units in the feature extraction phase.
2. 2. The removal of stop words is a strategy that keeps the overall context intact while excluding frequently used words with less significance from the text. Stop words belong to a class of words that are often employed in speech but are typically regarded as having little meaning or significance. Stop words contain phrases like "the," "is," and "and." The degree of interference may be successfully reduced by removing stop words from the content so that attention can be focused on finding more important phrases.
3. To calculate term frequency (TF), one must ascertain how frequently a specific term appears in a given document. By adding up the frequency of each term found in the tokenized text, the term frequency (TF) may be calculated. The more often occurring terms in a document, the more important they are.
4. One way to assess a term's infrequency within a corpus is to calculate its inverse document frequency (IDF). While terms with low frequency or occurrence in a small number of papers are seen to be more relevant, terms with high frequency throughout several documents are thought to be less important. The following formula may be used to calculate the IDF value: IDF(term) is calculated as log(N / (1 + DF(term)), where N is the total number of documents and DF(term) is the total number of documents containing the term.
5. DF(term) = the number of articles in which the word appears, while N is the total number of articles in the data, is the formula for determining the document frequency of a word (term) in a dataset.The computation of TF-IDF score involves the multiplication of term frequency (TF) and inverse document frequency (IDF) for each term. The TF-IDF metric quantifies the significance of a given term within a particular document as well as across the entire collection of documents. Terms that possess higher TF-IDF scores are deemed to be more representative of the content of the document and are regarded as more informative characteristics.
6. The process of feature representation involves the computation of TF-IDF scores for every term present in the document. The resulting feature vector serves as a representation of the document's content. Each individual term is associated with a particular feature, and its corresponding TF-IDF score represents the numerical value assigned to that particular feature. Subsequently, deep learning and machine learning models utilize the feature vector as their input.

To summarize, our methodology integrates various techniques including data preprocessing, TF-IDF feature extraction, a customized fully connected deep learning model, and a Random Forest model to identify indications of depression from text data on social media platforms. Our objective is to employ deep learning and machine learning methodologies to capture intricate patterns and characteristics that differentiate depressive text. The present study assesses the proposed methodology on a dataset that has been annotated, and subsequently compares its efficacy with pre-existing methodologies, thereby showcasing its proficiency in the classification of indications of depression through automation.

**Fully Connected Customized Deep Learning Model:** After performing feature extraction, the TF-IDF features are inputted into a customized deep learning model with fully connected layers, which is comparable to the previous methodology. The aforementioned model is capable of comprehending intricate patterns and interconnections present in the text, thereby enabling it to acquire significant representations of features associated with depression.

**Machine Learning Model:** Following the deep learning layers, a machine learning model, namely Random Forest, is introduced to enhance the classification process. The inputs to the Random Forest model are derived from the features extracted by the deep learning model. The Random Forest algorithm employs a collection of decision trees that form an ensemble, with each tree being trained on a distinct subset of the available data. The utilization of an ensemble approach allows for the model to produce resilient predictions through the consolidation of the outputs of discrete decision trees.

# Experiments

## Dataset Description

The social media post dataset utilized in our experiments was obtained from a publicly available repository. The corpus comprises textual excerpts extracted from social media platforms, with each excerpt being assigned a categorical label denoting the degree of depression severity. The dataset comprises three distinct classes, namely "not-depressed," "moderate," and "severe," which correspond to varying degrees of depressive symptoms.

Two separate files, a training sample file and a testing sample file, have been created from the provided dataset. The test set was used to evaluate the models' performance on new data while the training set was used to train and optimize the models. There were 8891 examples in the training data file, each with a text column and a label designating its class. The following was determined for the data distribution among the three classes: the "not-depressed" class consisted of 1971 examples, the "moderate" class consisted of 6019 examples, and the "severe" class consisted of 901 examples. The testing sample file comprised a total of 3245 samples, which were distributed across the classes as follows: 848 examples in the "not-depressed" class, 2169 examples in the "moderate" class, and 226 examples in the "severe" class.

The dataset furnished a heterogeneous collection of social media posts, affording us the opportunity to apprehend a spectrum of depression symptoms and gradations of severity. The objective of this study was to assess the efficacy of our proposed methodology for identifying indications of depression in social media text, using the provided dataset.

## Baseline Methodology

The baseline methodology employed in this study involved the utilization of a research investigation as a point of reference for the categorization of depression through the analysis of social media text. The study's foundation involved the implementation of a pre-existing BERT (Bidirectional Encoder Representations from Transformers) model that had undergone transfer learning methodologies.

* The BERT model, which is a transformer-based approach, has been pre-trained and has exhibited exceptional performance on a range of natural language processing tasks, such as text classification. The model has undergone pre-training on a substantial corpus of textual data, enabling it to proficiently capture contextual information and semantic relationships present within the text.
* Transfer learning is a methodology that utilises the expertise gained from pre-training on a vast dataset to enhance the efficacy on a specific task with a smaller dataset. The baseline investigation involved the customization of the pre-existing BERT model for the purpose of identifying depression through the analysis of social media language.
* Hyperparameter tuning is the process of determining the most suitable values for different parameters within a model with the aim of optimising its performance. The study's baseline methodology involved the utilisation of hyperparameter tuning techniques to identify the optimal configuration for the BERT model, thereby ensuring optimal performance in the classification of depression.
* The accuracy metric was utilised to evaluate the performance of the baseline methodology. The metric of accuracy pertains to the proportion of accurately classified instances in relation to the overall number of instances. The statement offers a comprehensive evaluation of the model's capacity to precisely forecast depression labels.
* 5. Results: According to the baseline study, the pre-trained BERT model with transfer learning achieved a maximum accuracy score of 57% for depression classification. The precision metric denotes the efficacy of the baseline methodology in accurately forecasting the existence or non-existence of depression through the analysis of social media textual data.

The efficacy of the baseline methodology can be ascribed to the robust functionalities of the pre-existing BERT model, which adeptly captures the contextual nuances and semantic associations present within the given text. The implementation of transfer learning methodologies has augmented the model's capacity to generalise and acclimatise to the particular undertaking of categorising depression. By means of hyperparameter tuning, the initial investigation attained the most optimal configuration for the BERT model, thereby enhancing its performance in the designated task.

To summarise, the fundamental approach utilised a pre-existing BERT model that had undergone transfer learning techniques to classify depression. The methodology exhibited encouraging outcomes, attaining a peak precision rate of 57%. The application of pre-existing models and transfer learning methodologies demonstrate the capacity to utilise prior knowledge and refine models for enhanced efficacy in the classification of depression.

## Evaluation Metrics

* In order to assess the efficacy of our proposed study in comparison to the baseline methodology, we employed the subsequent evaluation metrics:
* The metric of accuracy evaluates the general correctness of the model's predictions by computing the proportion of accurately classified instances to the total number of instances.
* Precision is a metric that evaluates the accuracy of positive predictions by determining the ratio of true positive samples to the total number of predicted positive samples. The emphasis is placed on the model's capacity to prevent the occurrence of false positives.
* The measure of recall, also referred to as sensitivity or true positive rate, pertains to the ratio of accurately predicted positive samples to the overall number of positive samples. The emphasis is placed on the model's capacity to prevent the occurrence of false negatives.
* The F1-score is a statistical measure that represents the harmonic mean of precision and recall. The metric offers an equitable evaluation of the model's efficacy, considering the occurrences of both type I and type II errors.

## Experimental Results

The study presents the implementation of four distinct machine learning models, namely K-Nearest Neighbours (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM), through a series of experiments. The metrics mentioned earlier were utilised to evaluate the efficacy of all of the models.

**K-Nearest Neighbors (KNN):** The K-Nearest Neighbours (KNN) model attained a precision rate of 0.52%. The classification report (Table 1) provides a summary of the precision, recall, and F1-score for each class. The K-Nearest Neighbours (KNN) model exhibited comparatively inferior performance in relation to the other models, thereby suggesting its limited ability to accurately classify indications of depression from textual data sourced from social media.

Table 1: Classification Report of KNN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report - KNN** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| Not Depressed | 0.68 | 0.66 | 0.67 | 2169 |
| Moderate | 0.29 | 0.29 | 0.29 | 848 |
| Severe | 0.09 | 0.12 | 0.1 | 228 |
|  |  |  |  |  |
| accuracy |  |  | 0.52 | 3245 |
| macro avg | 0.35 | 0.35 | 0.35 | 3245 |
| weighted avg | 0.53 | 0.52 | 0.53 | 3245 |

**Decision Tree:** The Decision Tree model attained a precision rate of 0.58%. Table 2 displays the precision, recall, and F1-score metrics for each class in the classification report. Although the Decision Tree model outperformed the KNN model, its ability to accurately capture the subtleties of depression indicators in social media text remained restricted.

Table 2: Classification Report of Decision Tree

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report – Decision Tree** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| Not Depressed | 0.68 | 0.76 | 0.72 | 2169 |
| Moderate | 0.33 | 0.24 | 0.28 | 848 |
| Severe | 0.10 | 0.09 | 0.1 | 228 |
|  |  |  |  |  |
| accuracy |  |  | 0.58 | 3245 |
| macro avg | 0.37 | 0.36 | 0.36 | 3245 |
| weighted avg | 0.55 | 0.58 | 0.56 | 3245 |

**Random Forest:** The Random Forest algorithm exhibited superior performance compared to the preceding models, attaining an accuracy score of 0.65%. The classification report (Table 3) presents the precision, recall, and F1-score for each class. The Random Forest model exhibited enhanced performance, indicating its proficiency in capturing intricate associations within the textual data and generating more precise prognostications for depression categorization.

Table 3:Classification Report of Random Forest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report – Random Forest** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| Not Depressed | 0.68 | 0.90 | 0.78 | 2169 |
| Moderate | 0.39 | 0.16 | 0.23 | 848 |
| Severe | 0.27 | 0.01 | 0.03 | 228 |
|  |  |  |  |  |
| accuracy |  |  | 0.65 | 3245 |
| macro avg | 0.45 | 0.36 | 0.34 | 3245 |
| weighted avg | 0.58 | 0.65 | 0.58 | 3245 |

**Support Vector Machine (SVM):** The support vector machine (SVM) model demonstrated superior performance compared to other machine learning models, exhibiting an accuracy score of 0.67%. The classification report (Table 4) contains the precision, recall, and F1-score for each class. The Support Vector Machine (SVM) model demonstrated favourable performance in precisely categorising indications of depression from textual data on social media, thereby presenting additional proof of its efficacy.

Table 4:Classification Report of SVM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report - SVM** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| Not Depressed | 0.67 | 1.00 | 0.80 | 2169 |
| Moderate | 0.00 | 0.00 | 0.00 | 848 |
| Severe | 0.00 | 0.00 | 0.00 | 228 |
|  |  |  |  |  |
| accuracy |  |  | 0.67 | 3245 |
| macro avg | 0.22 | 0.33 | 0.27 | 3245 |
| weighted avg | 0.45 | 0.67 | 0.54 | 3245 |

**Comparison with Baseline Study:** Compared to the baseline methodology, our proposed study incorporating the deep learning model followed by Random Forest yielded improved results. The baseline study achieved a maximum accuracy score of 57%, while our proposed study achieved higher accuracy scores ranging from 0.52% to 0.67% for different models (Table #).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Results Comparison** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| Baseline (Deproberta) | 0.55 | 0.43 | 0.48 | 3245 |
| DL + KNN | 0.53 | 0.52 | 0.53 | 3245 |
| DL + Decision Tree | 0.55 | 0.58 | 0.56 | 3245 |
| DL + RF | 0.58 | 0.65 | 0.58 | 3245 |
| DL + SVM | 0.45 | 0.67 | 0.54 | 3245 |

The Random Forest algorithm exhibited superior performance compared to the baseline study with respect to metrics such as accuracy, precision, recall, and F1 score. The aforementioned statement highlights the efficacy of our suggested methodology in precisely identifying indications of depression from textual content on social media platforms. Furthermore, it outperforms the baseline approach in terms of performance.

The implementation of machine learning models, in conjunction with the feature extraction methodology utilising TF-IDF, yielded significant insights into textual features associated with indicators of depression. Consequently, this has resulted in an enhancement of the efficacy of our models. In brief, our empirical investigations involving diverse machine learning models have revealed that the Random Forest model exhibits the greatest resilience in accurately classifying textual data for the purpose of detecting depression.

# Conclusion

The mental illness known as depression is widely prevalent and can have significant negative impacts on the overall well-being and quality of life of those affected. The identification and categorization of indications of depression through social media language can offer significant perspectives for prompt intervention and assistance. The objective of this investigation was to tackle this challenge through the presentation of a multi-classification methodology that employs BERT models for the identification of indications of depression. Our research endeavours and subsequent analyses have yielded noteworthy advancements in the realm of depression categorization based on social media language. Our approach involved the implementation of TF-IDF for the purpose of feature extraction. This enabled us to identify and capture significant textual features that are associated with indicators of depression. The aforementioned methodology facilitated the extraction of significant insights from the textual data and furnished valuable inputs to our models.

The approach that was proposed involved utilising a tailored deep learning model in conjunction with machine learning models such as K-Nearest Neighbours, Decision Tree, Random Forest, and Support Vector Machine. The integration of deep learning and conventional machine learning techniques facilitated the exploitation of their respective advantages. The deep learning model demonstrated proficiency in capturing intricate patterns and interconnections within the textual data, whereas the machine learning models leveraged the extracted features to achieve precise predictions for depression classification. Our study involved a comparative analysis between our findings and a baseline approach that utilised a pre-trained BERT model with transfer learning. The approach we proposed exhibited exceptional performance in accuracy, precision, recall, and F1-score metrics across all machine learning models. The Random Forest Machine model demonstrated superior accuracy scores, indicating its efficacy in precisely identifying indications of depression in social media text.

The obtained outcomes hold considerable significance for various reasons. The results of our study underscore the promise of employing machine learning methodologies, in tandem with deep learning and feature extraction, for the purpose of accurately categorising indications of depression from textual data on social media. The aforementioned has pragmatic ramifications for mental health practitioners and scholars, who can utilise these techniques for prompt detection and remediation. In addition, our research adds to the expanding corpus of literature on identifying depression through the analysis of social media information. Our proposed methodology involves the integration of TF-IDF with a hybrid ensemble of deep learning and machine learning models, thereby presenting a supplementary strategy that complements conventional methodologies. The integration of these methodologies yields a sturdy and dependable structure for the categorization of depression, augmenting the precision and efficacy of the categorization procedure.

Nevertheless, it is crucial to acknowledge certain constraints of our research. The efficacy of our models is significantly contingent upon the calibre and inclusivity of the training data. Acquiring a varied and all-encompassing dataset that precisely reflects the social media text of the intended populace is imperative in enhancing the efficacy of the model. Moreover, there is a need for additional investigation to examine alternative feature extraction methodologies and model structures that can augment the classification precision and extend the results to diverse social media channels.

In summary, the research offers a thorough and efficient methodology for identifying indications of depression through textual content on social media platforms. Through the integration of TF-IDF, deep learning, and machine learning models, we were able to attain a superior level of performance in comparison to the baseline methodology. The findings of our research provide significant perspectives on the advancement of mechanised instruments for prompt identification and interference in the realm of mental health. It is imperative to further investigate and enhance these methodologies in order to tackle the changing complexities in the categorization of depression and to provide assistance to individuals who may be encountering symptoms of depression.

# References:

De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E., 2013. Predicting Depression via Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, [online] 7(1), pp.128–137. https://doi.org/10.1609/ICWSM.V7I1.14432.

Coppersmith, G., Dredze, M. and Harman, C., 2014. Quantifying Mental Health Signals in Twitter. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, [online] pp.51–60. https://doi.org/10.3115/V1/W14-3207.

Desmet, B. and Hoste, V., 2013. Emotion detection in suicide notes. *Expert Systems with Applications*, [online] 40(16), pp.6351–6358. https://doi.org/10.1016/J.ESWA.2013.05.050.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, [online] 1, pp.4171–4186. Available at: <https://arxiv.org/abs/1810.04805v2> [Accessed 12 June 2023].

Guntuku, S.C., Preotiuc-Pietro, D., Eichstaedt, J.C. and Ungar, L.H., 2019. What Twitter Profile and Posted Images Reveal about Depression and Anxiety. *Proceedings of the International AAAI Conference on Web and Social Media*, [online] 13, pp.236–246. https://doi.org/10.1609/ICWSM.V13I01.3225.

Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H. and Eichstaedt, J.C., 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18, pp.43–49. https://doi.org/10.1016/J.COBEHA.2017.07.005.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J., 2013. Distributed Representations of Words and Phrases and their Compositionality. [online] Available at: <http://arxiv.org/abs/1310.4546> [Accessed 12 June 2023].

Pennington, J., Socher, R. and Manning, C.D., 2014. GloVe: Global Vectors for Word Representation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.1532–1543. https://doi.org/10.3115/V1/D14-1162.