
MENTAL HEALTH CONDITION MONITORING FROM SOCIAL MEDIA

DATA SCIENCE IN AI FOR HEALTHCARE PROJECT

CSE, 6TH SEMESTER

Group Members:

KARIMI UPENDRA	B121023
KETHAVATH RAJA NAYAK	B121024
THUTTA JAHNAVI	B121067

Submitted to:

DR. SANJAY SAXENA



Department of Computer Science and Engineering

International Institute of Information Technology

Bhubaneswar – 751003

Date: 25th April 2024

Contents

1. Abstract	3
2. Project Description	4
2.1 Project Title	4
2.2 Project Objective	4
2.3 Introduction	4
3. Materials and methods.....	5
3.1 Dataset	5
3.2 Preprocessing.....	6
3.3 Word Embedding.....	6
3.3.1 TFIDF(Term Frequency and Inverse Document Frequency).....	6
3.3.2 Bag of Words(BoW)	7
3.4 Traditional ML Models	7
3.5 Deep Learning Model.....	8
3.6 Evaluation Metrics.....	8
4. Experimental Results.....	9
4.1 Word Cloud	9
5. Discussion	10
6. Conclusion.....	10
References.....	11

1.ABSTRACT

Mental health detection from social media platforms has emerged as a vital area of research, offering insights into individuals' psychological well-being and facilitating early intervention. In this study, we explore the task of detecting mental health-related content from Reddit posts. Leveraging a dataset curated by Turcan and McKeown, we employed both traditional machine learning algorithms and deep learning techniques to develop accurate classifiers.

Our experiments revealed promising results across various methodologies. Among traditional machine learning algorithms, logistic regression and support vector machines (SVM) demonstrated competitive performances, achieving accuracies of 74.83% and 73.85%, respectively. These findings underscore the efficacy of conventional approaches in handling textual data for mental health detection tasks.

Furthermore, we investigated the application of deep learning architectures, specifically Long Short-Term Memory (LSTM) networks, renowned for their ability to capture sequential dependencies in data. Remarkably, the LSTM model outperformed traditional methods, achieving an accuracy of 78.01%. This underscores the efficacy of deep learning approaches in capturing nuanced linguistic patterns indicative of mental health concerns.

Our study contributes to the growing body of research aimed at leveraging computational methods for mental health analysis. The findings suggest that both traditional and deep learning techniques hold promise for automating the detection of mental health-related content on social media platforms, thereby facilitating timely interventions and support for individuals in need.

2. Project Description

2.1. Project Title

MENTAL HEALTH CONDITION MONITORING FROM SOCIAL MEDIA

2.2. Project Objective

The primary objective of this project is to develop a predictive model to assess the mental health status of individuals based on natural language input, such as text data. By analysing linguistic patterns, sentiment, and other relevant features, the model aims to provide valuable insights into an individual's mental well-being and predict their emotional state. Predicting the emotional state of an individual can help enhance mental health support by providing valuable insights into their current well-being. By anticipating emotions, our goal is to support people in taking proactive measures to care for their mental well-being, leading to enhanced quality of life and a stronger emotional foundation.

2.3. Introduction

In today's fast-paced world, stress, anxiety, and burnout are increasingly common. Individuals who suffer from anxiety and depression frequently express their views and ideas on social media. Thus, several studies found that people who are contemplating stress can be identified by analysing social media posts. However, finding and comprehending patterns of stress represent a challenging task. Therefore, it is essential to develop a machine learning system for automated early detection of depression or any abrupt changes in a user's behaviour by analysing his or her posts on social media. In this report, we propose a methodology based on experimental research for building a stress detection system using publicly available Reddit datasets, word-embedding approaches, such as TF-IDF and Word2Vec, for text representation, and hybrid deep learning and machine learning algorithms for classification.

Research shows that promoting happiness and well-being can lead to numerous benefits, including improved mental health, better relationships, and increased productivity. The project aims to address this need by offering practical guidance, motivational content, and a sense of community to support individuals in their journey toward greater happiness and fulfilment.

Keywords: Mental health, Social Media, Natural Language Processing, Machine Learning, Deep Learning, Reddit posts, TF-IDF, Bag of Words, Logistic Regression, Support Vector Machines, LSTM, Classification, Preprocessing, Word embedding.

3. Materials and Methods

3.1. Data

In this study, we acquired data from multiple sources to construct a comprehensive dataset for the task of detecting mental health-related content from Reddit posts. Initially, we examined four research papers to identify potential datasets suitable for our objectives. Ultimately, the dataset obtained from the research paper authored by Elsbeth Turcan and Kathleen McKeown was selected for its relevance and suitability.

The dataset provided by **Turcan and McKeown** consisted of *3000 rows* and *116 columns*, containing a wealth of information extracted from Reddit posts. These features encompassed various textual, temporal, and user-related attributes, offering a rich foundation for our analysis.

To augment our dataset and enhance its diversity, we explored additional sources, including Kaggle. While we discovered another dataset on Kaggle, it lacked labels for the posts, presenting a challenge for supervised learning tasks. To address this limitation, we employed **K-means** clustering to automatically categorize the posts based on their textual content. Subsequently, we manually verified the clustering results for approximately *100-150 posts* to ensure accuracy.

The manual verification process involved assessing the assigned labels against the content of the posts to ascertain their alignment with mental health-related themes. Through this meticulous validation process, we verified the correctness of the assigned labels, thereby enhancing the reliability of our dataset.

Additionally, while reviewing the research papers, we found a plethora of potential information relevant to our study. Insights gleaned from these papers not only informed our dataset curation process but also provided valuable context for understanding the nuances of mental health discourse on social media platforms.

By integrating data from multiple sources, employing rigorous labelling approaches, and leveraging insights from existing research, we curated a robust dataset tailored to our specific research objectives. This comprehensive dataset served as the foundation for training and evaluating our machine learning models, enabling us to effectively address the task of mental health detection from Reddit posts.

3.2.Preprocessing

We focused our analysis on a subset of the dataset, specifically, two columns out of the 116 available. Those two columns include only the text of the social media post and a binary label for that post telling whether the post implies stressed state or not. Data Preprocessing is done using the techniques of Natural Language Processing. Our preprocessing pipeline, outlined in the code snippet provided, entailed the following steps:

- **Lowercasing:** Converting all text to lowercase to ensure uniformity.
- **URL Removal:** Eliminating any URLs present in the text to remove irrelevant information. **HTML Tag Removal:** Stripping off HTML tags, as they do not contribute to the textual content.
- **Digits and Single Characters Removal:** Omitting digits and single characters to enhance readability and focus on meaningful words.
- **Punctuation Removal:** Discarding punctuation marks to isolate words for analysis. **Tokenization and Stop word Removal:** Splitting the text into individual words (tokens) and removing common English stop words, such as "the," "is," and "and," to eliminate noise. **Word Stemming:** Applying the Porter Stemmer algorithm to reduce words to their root form, thereby consolidating similar words and reducing redundancy.
- **Recomposition:** Reassembling the processed words into a coherent, cleaned-up text.

3.2.Word Embedding

3.2.1. TFIDF

TF-IDF is a widely used technique in natural language processing (NLP) for converting textual data into numerical vectors. It represents the importance of a word in a document relative to a corpus of documents. TF-IDF is calculated as the product of two components: Term Frequency (TF) and Inverse Document Frequency (IDF).

$$TF(w)_d = \frac{n_w(d)}{|d|} \quad (1)$$

Set D points to a set of documents, and d denotes a single document, $d \in D$. Each document is represented as a group of sentences and words w , and $n_w(d)$ is the number of recurrent words w in document d . Therefore, the size of document d is calculated as follows:

$$|d| = \sum_{w \in d} n_w(d) \quad (2)$$

The frequency at which a word appears in the document is expressed in Equation (2). *IDF*, the second component of *TF-IDF*, is used to compute the number of documents in a textual corpus in which a specific word appears, as follows:

$$IDF(w)_d = 1 + \log\left(\frac{|D|}{|\{d : w \in d\}|}\right) \quad (3)$$

The *TF-IDF* for word w associated with document d and corpus D can be calculated as:

$$TF - IDF = TF(w)_d \times IDF(w)_D \quad (4)$$

3.2.2. Bag of Words (BoW)

The Bag of Words (BoW) model is a fundamental technique in Natural Language Processing (NLP) for converting textual data into numerical vectors. It represents the occurrence of words in a document without considering the order in which they appear.

Tokenization: The first step in the Bag of Words model involves tokenization, where the text is split into individual words or tokens.

- **Vocabulary Creation:** Next, a vocabulary is created by collecting all unique words (or tokens) present in the corpus of documents.
- **Vectorization:** Each document is represented as a vector, where each dimension corresponds to a word in the vocabulary. The value of each dimension represents the frequency of the corresponding word in the document.
- **Example:** Consider the following two sentences: "The cat sat on the mat." and "The dog played in the yard." The vocabulary would consist of ["the", "cat", "sat", "on", "mat", "dog", "played", "in", "yard"]. The vector representations of these sentences would be [1, 1, 1, 1, 1, 0, 0, 0, 0] and [1, 0, 0, 0, 0, 1, 1, 1, 1], respectively.

3.3. Traditional Machine Learning Models

After employing TF-IDF for word embedding, we proceeded to train seven machine learning models to classify Reddit posts into mental health-related categories. These models were chosen based on their versatility and effectiveness in handling text data. The models utilized are as follows:

1. Logistic Regression
2. KNN Classifier (K-Nearest Neighbors)
3. Random Forest Classifier
4. Decision Tree Classifier
5. Naive Bayes Classifier
6. AdaBoost Classifier
7. SVM Classifier (Support Vector Machine)

Following rigorous training and evaluation procedures, it was observed that the Logistic Regression and SVM classifiers exhibited the highest accuracy among all models. These two algorithms demonstrated superior performance in discerning patterns and predicting mental health conditions based on social media content.

Among the seven models evaluated, Logistic Regression and SVM classifiers demonstrated the highest accuracy in predicting the mental health-related categories of Reddit posts with 74.83% and 73.85% This indicates the effectiveness of these models in capturing the underlying patterns in the TF-IDF weighted features.

3.4.Deep Learning Model(LSTM)

LSTM networks are a variant of recurrent neural networks (RNNs) designed to overcome the vanishing gradient problem and effectively model long-range dependencies in sequential data. They are particularly well-suited for tasks involving sequential or time-series data, such as natural language processing. Our LSTM model was trained using the Reddit post data represented in the Bag of Words format. The BoW representation captures the occurrence of words in a document, providing a numerical input suitable for deep learning models. During training, the model learned to map the BoW representations to the corresponding mental health categories.

Table 3.4.1: Parameters and their values used in LSTM model.

Parameter	Value
Input Sequence Length	142
Embedding Dimension	100
Vocabulary Size	7823
LSTM Units	128
Dropout	0.5
Batch Size	64
Number of Epochs	10
Activation Function	'sigmoid'

3.4.Evaluation Metrics

To evaluate the performance of the models in classifying post content as suicidal or non-suicidal, we used common evaluation metrics with a focus on the number of false-positive and false-negative classifications obtained from the confusion matrix presented. The performance metrics used were *Accuracy*, *Precision*, *Recall*, and *Standard Deviation*, which were calculated as follows:

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \times 100$$

$$Precision = \frac{TP}{TP + FP} \times 100$$

$$Recall = \frac{TP}{TP + FN} \times 100$$

4.Experimental Results

The LSTM model achieved an accuracy of 78.01% using bag of words and Logistic regression model got an accuracy of 74.83 using TF-IDF word embedding. The followings results were obtained:

Table 4.1: Results of traditional ML models

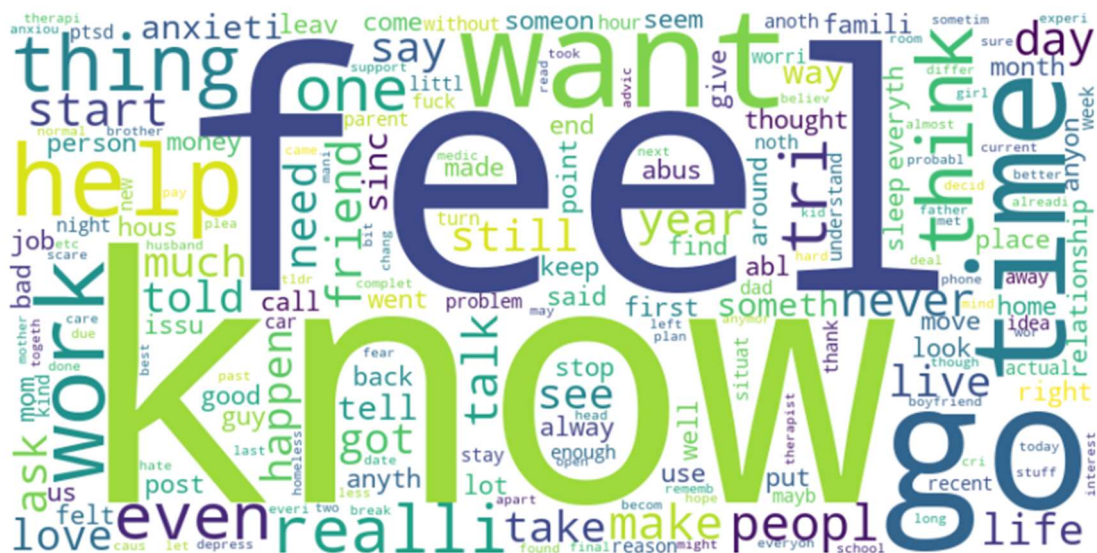
Classifier	Accuracy	Precision	Recall	Standard Deviation
Logistic Regression	74.83	79.95	73.57	0.030
KNN	64.48	72.36	63.72	0.20
Random Forest	68.95	82.11	66.01	0.036
Decision Tree	60.28	46.61	66.41	0.021
Naïve Bayes	54.55	68.56	54.76	0.020
Adaboost	66.99	69.11	67.64	0.022
SVM	73.85	77.51	73.33	0.006

Table 4.2: Model results

Model	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
LSTM	0.9965	0.0165	0.7801	0.9163

4.1. Word Cloud

Word clouds are widely used in NLP to visualize the most important and recurrent words in a textual corpus. Here, we used a word cloud to visualize the most repeated words in the Reddit dataset, shown below.



5. Discussion

The project focused on detecting mental health-related content from Reddit posts using machine learning and deep learning techniques. By leveraging a diverse dataset and employing various models such as logistic regression, support vector machines, and LSTM networks, the study aimed to automate the identification of mental health discussions on social media platforms.

The findings of the project highlighted the effectiveness of both traditional machine learning algorithms and deep learning architectures in accurately classifying mental health-related content. The logistic regression and SVM models demonstrated competitive performance, while the LSTM model showcased the ability to capture nuanced linguistic patterns inherent in mental health discussions.

6. Conclusion

In conclusion, this project signifies the pivotal role of machine learning and deep learning techniques in automating the detection of mental health-related content on social media platforms. By leveraging diverse datasets and employing a range of models, including logistic regression, support vector machines, and LSTM networks, the study showcases promising results in accurately identifying and flagging mental health discussions. The findings not only highlight the effectiveness of these computational methods but also underscore their potential impact in facilitating early interventions, destigmatizing mental health discourse, and ultimately improving the well-being of individuals in online communities. This research contributes to the ongoing efforts to harness technology for addressing societal challenges, particularly in the domain of mental health, and underscores the importance of interdisciplinary collaborations between computer science, psychology, and healthcare domains.

References

Multitask Learning for Mental Health: Depression, Anxiety, Stress (DAS) Using Wearables
by Berrenur Saylam and Özlem Durmaz İnel.

Link <https://www.mdpi.com/2075-4418/14/5/501>

Multi-task learning to detect suicide ideation and mental disorders among social media users
by Prasadith Buddhitha and Diana Inkpen.

Link: <https://www.frontiersin.org/articles/10.3389/frma.2023.1152535/full>

CLPsych2019 Shared Task: Predicting Users' Suicide Risk Levels from Their Reddit Posts
on Multiple Forums Victor Ruiza, Lingyun Shia, Jorge Guerra^{a,b}, Wei Quana^c, Neal Ryand,
Candice Biernesser^d, David Brent^d, and Fuchiang Tsui^{a-c}

Link: https://drive.google.com/file/d/1o6E92ceOF7fn6_Dxn8nJ3eP5c_CwI34o/view

Dreaddit: A Reddit Dataset for Stress Analysis in Social Media Elsbeth Turcan, Kathleen
McKeown Columbia University Department of Computer Science

Link: <https://arxiv.org/pdf/1911.00133v1.pdf>