

# Plagiarism - Report

Originality Assessment

5%



Overall Similarity

**Date:** Mar 15, 2024

**Matches:** 120 / 2353 words

**Sources:** 4

**Remarks:** Low similarity detected, check with your supervisor if changes are required.

**Verify Report:**

## ABSTRACT

Detecting depressive symptoms is crucial for improving mental health outcomes, especially when dealing with a vast textual dataset like social media posts. In our study, we introduced a predictive method leveraging Naive Bayes, which demonstrated promising accuracy in identifying depressive symptoms. These findings underscore the potential utility of Naive Bayes for early depression detection within large populations. Our method begins by extracting various text features, including word usage patterns, sentiment analysis, and text structure analysis. These features serve as input to train a Naive Bayes classifier. This classifier learns the likelihood of each feature occurring in texts associated with depression and those that are not. Subsequently, it employs these learned probabilities to make predictions about whether a given text exhibits depressive symptoms or not.

Keywords: Depressive Symptoms, Depression, Machine Learning

## 1. INTRODUCTION

### 1.1 DEPRESSIVE SYMPTOMS

Depression is a multifaceted and widespread mental health disorder that affects millions of individuals worldwide. This condition is characterized by a diverse array of symptoms related to low mood, which can vary in both intensity and duration. These symptoms exert a profound impact on an individual's emotional state, cognitive abilities, and physical health. Grasping the breadth of depressive

symptoms and their significance is crucial for the identification and effective management of this condition. In this discussion, we will delve into the different manifestations of depressive symptoms, their consequences, and the importance of seeking assistance when necessary. It's essential to understand that **2 depression is a** treatable condition, and with the appropriate support and interventions, individuals can recover their mental and emotional well-being.

### 1.2 DEPRESSION

**Depression is a** widespread and debilitating mental health disorder that transcends various

demographic factors such as age, gender, and cultural background. It's characterized by persistent and overwhelming feelings of sadness, hopelessness, and a pervasive loss of interest in enjoyable activities. While occasional sadness is a normal human experience, depression goes much deeper, significantly affecting daily functioning, relationships, and overall well-being. This complex condition presents itself through a range of emotional and physical symptoms, and it often carries a substantial societal stigma. However, it's essential to acknowledge and comprehend depression because of its profound impact on individuals and their social circles.

## 2. LITERATURE REVIEW

### 2.1 DEPRESSION AND OTHER COMMON MENTAL DISORDERS

The [2 Global Health Estimates](#), along with other relevant sources, have underscored the vital importance of having reliable and up-to-date data on the prevalence of various diseases and health conditions within a general population. Such data forms a critical foundation for the development and implementation of effective health policies, strategic planning, and outcome evaluation. Over time, epidemiological studies have been conducted in diverse low-, middle-, and high-income countries, significantly enhancing our understanding of the global prevalence of psychiatric disorders. These studies have contributed substantially to our knowledge of the worldwide [2 prevalence and incidence](#) rates of psychiatric disorders. Additionally, there have been notable advancements in quantifying both fatal and non-fatal outcomes associated with these disorders, including the measurement of time spent in a state of ill health.

### 2.2 JUMPING NLP CURVES: A REVIEW OF NATURAL LANGUAGE PROCESSING RESEARCH

Erik Cambria and his colleagues have introduced a system that highlights the evolution of Natural Language Processing (NLP) as a field. NLP is a set of computational techniques aimed at the automatic analysis and representation of human language. The journey of NLP research has transformed significantly over the years, transitioning from the era of punch cards and batch processing, where analyzing a sentence could take up to 7 minutes, to the contemporary era represented by technology giants like Google, where millions of Webpages can be processed in less than a second. This review paper sheds light on recent developments in NLP research, offering a fresh

perspective on the historical progression, the current state, and the potential future of NLP technology.

### 2.3 AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS

Erik Cambria and his colleagues have proposed a system that emphasizes the significance of understanding emotions in the realm of personal development and human growth. This understanding is considered a fundamental component in the pursuit of replicating human intelligence. Beyond its importance in advancing artificial intelligence (AI), the processing of emotions holds **1 a crucial role** in tasks closely related to polarity detection. **2 The ability to** automatically gauge public sentiments concerning social events, political movements, marketing campaigns, and product preferences have garnered considerable interest. This interest is observed both within the scientific community, where it presents exciting challenges, and in the business world, where it offers substantial benefits in marketing and financial market prediction. This trend has given rise to emerging fields such as affective computing and sentiment analysis.

### 2.4 CAPTURING DEPENDENCIES AMONG LABELS AND FEATURES FOR MULTIPLE EMOTIONS TAGGING OF MULTIMEDIA

In their research, Shan Wu and colleagues addressed the challenge of emotion tagging for multimedia data. They aimed to capture the interdependencies among multiple emotions in both the features extracted from **1 the data and the** emotion labels assigned to that data. These dependencies are crucial for enhancing the analysis of affective content in multimedia but have not been fully explored in previous work. To tackle this issue, the researchers introduced two hierarchical models. These models were designed to learn shared features and global semantic relationships among emotion labels, enabling them to jointly assign multiple emotion labels to multimedia data. Additionally, the researchers developed efficient algorithms for learning and inference in these models. Their experiments, conducted on three benchmark emotion databases, demonstrated that their proposed methods outperformed existing approaches.

### 2.5 A HYBRID MODEL FOR AUTOMATIC EMOTION RECOGNITION IN SUICIDE NOTES

In their system, Hui Yang et al. presented their submission to the 2011 i2b2/VA/Cincinnati Medical Natural Language Processing Challenge, Track 2 Shared Task, which focused on sentiment analysis in suicide notes at the sentence level. Their approach is a hybrid model that incorporates various natural

language processing techniques, including lexicon-based keyword spotting, CRF-based emotion cue identification, and machine learning-based emotion classification. To consolidate the results from these techniques, they employed different vote-based merging strategies. Remarkably, their automated system demonstrated strong performance when compared to the manually-annotated gold standard. It achieved highly promising results with a micro-averaged F-measure score of 61.39% in textual emotion recognition.

Top of Form

### 3. EXISTING SYSTEM

**3** Major depressive disorder (MDD) is a mental health problem that is widely widespread and has emerged as a significant social challenge. Early treatment response prediction may be helpful in depression rehabilitation engineering, which is extremely important **2** in terms of easing the burden and suffering associated with MDD. In this research, we describe a sequence modeling strategy that predicts individuals whose response to therapy is indicated by a decrease in clinically administered scales. The data obtained through passive sensing techniques **1** is used to make this prediction. Four psychiatric sites' outpatient clinics have yielded hundreds of MDD patients. A self-developed app has been provided for each user to passively record their daily phone usage and physical data with the least amount of human intervention. Missing values in passive sensing is an inevitable problem. **1** In order to get around that, the suggested strategy integrated techniques for feature extraction and sequence modeling to make full use of the pattern of missing values from longitudinal data.

### 4. PROPOSED SYSTEM

The proposed system represents a straightforward yet highly effective approach for the prediction of depressive symptoms within extensive textual datasets. Leveraging the well-established Naive Bayes machine learning algorithm, this system capitalizes on a substantial collection of social media posts, a valuable resource for understanding depression. Its potential significance lies in its capacity to serve as a valuable tool for early depression detection, which is pivotal for enhancing depression-related outcomes. Early detection of depression bears substantial importance in improving overall mental health outcomes. The proposed system demonstrates promise in identifying individuals at risk of developing depression, while also serving **3** as a means to monitor the progress of those already

undergoing treatment. Its primary function involves the analysis of textual data to categorize individuals into two groups: those exhibiting **2 depressive symptoms and** those who do not. To optimize accuracy and comprehensibility, the system incorporates feature selection techniques and draws upon domain-specific knowledge.

#### 4.1 LOAD INPUT DATA

During this stage, we compile and transfer the unprocessed text-based information into our analytical platform. This information usually comprises content generated by users, including posts on social media, interactions on forums, or feedback from surveys. This content may contain pertinent details related to **3 symptoms associated with** depression.

#### 4.2 DATA PREPROCESSING

**4 Data preprocessing is a critical step in** preparing text **data for analysis.** It encompasses various tasks aimed at enhancing the data's suitability and quality. These tasks involve breaking text into tokens (tokenization), eliminating extraneous information such as stop words and special characters, and standardizing text by converting it to lowercase. These procedures contribute to data consistency and overall data quality improvement.

#### 4.3 FEATURE SELECTION

Feature selection **1 plays a crucial role in** preparing preprocessed text data for **machine learning algorithms.** When working with the Naive Bayes (NB) algorithm, it's essential to convert text data into numerical features that the algorithm can effectively utilize. One common approach to achieve this is by quantifying the frequency of words or phrases within the text. Feature selection serves as a powerful tool **1 to enhance the** data preprocessing stage. It aids in the reduction of dimensionality, which, in turn, **2 can lead to** improved model performance. By carefully selecting the most **1 relevant features from the** text data, we can streamline the input variables and enhance the algorithm's ability to extract meaningful patterns and make accurate predictions.

#### 4.4 TRAINING AND TESTING

During this stage, the dataset **is divided into** two distinct subsets: a training set and a testing set. The training set **plays a crucial role in** the process as it serves as the foundation for instructing a machine learning model, like Naive Bayes, to discern underlying patterns within the data that pertain to

depressive symptoms. Conversely, the testing set is employed to evaluate the Model's proficiency in making predictions on fresh, unobserved data.

#### 4.5 PERFORMANCE EVALUATION

The performance of the NB model is evaluated using a variety of metrics, such as accuracy, precision, recall, and F1 score. The performance **1 of the model** can be improved by adjusting the hyperparameters **of the model** or by using a different feature selection method.

#### 5. RESULT ANALYSIS

algorithm

Precision

Recall

F score

Accuracy

NB

0.73

0.7

0.74

0.81

SVM

0.8

0.81

0.84

0.1

Table 1. Comparison table

Figure 1. Comparison graph

The results show different performance metrics when comparing the effectiveness of two classification algorithms, Naive Bayes (NB) and Support Vector Machine (SVM), for the job of identifying depressed symptoms. With a precision of 0.73, recall of 0.7, F score of 0.74, and accuracy of 0.81, Naive Bayes was shown to be effective. This shows that Naive Bayes produces a relatively high overall accuracy while displaying a balanced performance **2 in terms of** precision and recall. Conversely, the SVM method yielded an elevated F score of 0.84 because to its higher precision (0.8) and recall (0.81). SVM's accuracy was noticeably reduced at 0.1, though. Together, these metrics highlight the trade-offs between each algorithm's precision, recall, and overall accuracy, shedding light on how well the algorithms identify depressed symptoms in the dataset under study.

## 7. CONCLUSION

In summary, the advancement **1 of machine learning** models aimed at predicting depressive symptoms through the analysis of textual data represents a crucial stride in harnessing technology to augment mental health support and early intervention efforts. The delineated processes encompassing data collection, preprocessing, feature selection, training, and evaluation, in conjunction with well-considered input and output interface design, provide a solid groundwork for the development of resilient systems. The Naive Bayes (NB) algorithm emerges as a promising instrument for enhancing the early detection and **2 treatment of depression**. It stands out for its simplicity, efficiency, and reasonable accuracy, offering versatile applications to assist individuals grappling with depression.



Nevertheless, there remain several hurdles that must be surmounted before widespread deployment of the NB algorithm can be realized.

## 8. FUTURE WORK

Enhancing the feature extraction methods for detecting **2 depression is a** pressing area for advancement. Presently, state-of-the-art techniques rely heavily on word counts or TF-IDF, which may fall short in capturing the intricate subtleties of language pertinent to depression. Prospective research avenues should strive to devise more robust feature extraction approaches capable of better encapsulating these nuances. **3 In addition to** refining feature extraction, **there is a** compelling need to consider more sophisticated machine learning algorithms. While Naive Bayes (NB) serves as a straightforward and efficient choice, its predictive accuracy for depressive symptoms might not be optimal. Future investigations could delve into the application of advanced **1 machine learning algorithms**, such as deep learning models, **to enhance the** precision of depression predictions.

## 9. REFERENCES

1. World Health Organization (2020). **2 "Global Health Estimates** for Depression and Other Common Mental Disorders" (No. WHO/MSD/MER/2017.2).
2. World Health Organization (2021). **"Global Health Estimates** for Depression and Other Common Mental Disorders" (No. WHO/MSD/MER/2017.2).
3. Cambria, E. (2021). "Affective Computing and Sentiment Analysis." IEEE Intelligent Systems, 31(2), 102–107.
4. Shangfei, W., Guozhu, P., Zhuangqiang, Z., & Zhiwei, Xu. (2021). "Capturing Emotion Distribution for Multimedia Emotion Tagging." IEEE Transactions on Affective Computing.
5. Yang, H., Willis, A., De Roeck, A., & Nuseibeh, B. (2021). "A Hybrid Model for Automatic Emotion Recognition in Suicide Notes." Biomedical Informatics Insights, 5(1), 17–30.
6. Allouch M, Azaria A, Azoulay R, Ben-Izchak E, Zwilling M, Zachor DA (2021) Automatic detection of insulting sentences in conversation. Paper presented at: Proceedings of the 2021 IEEE international conference on the science of electrical engineering in Israel (ICSEE); pp 1–4; IEEE.
7. .Ding Y, Chen X, Fu Q, Zhong S (2020) A depression recognition method for college students using deep integrated support vector algorithm. IEEE Access 8:75616–75629

8. Uddin MZ, Torresen J (2019) Activity recognition using smartphone sensors, robust features, and recurrent neural network. In: 2019 13th international symposium on medical information and communication technology (ISMICT), May 2021

9. Kumari T, Mishra A (2020) Deep networks and deep learning algorithms. In: Artificial intelligence trends for data analytics using machine learning and deep learning approaches, pp 199–214

Uddin MZ, Hassan M, Alsanad A, Savaglio C (2020) A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare Inform Fusion

<https://doi.org/10.1016/j.inffus.2019.08.004>

## Sources

1	<a href="http://www.geeksforgeeks.org/data-preprocessing-in-data-mining/">www.geeksforgeeks.org/data-preprocessing-in-data-mining/</a> INTERNET 2%
2	<a href="https://www.who.int/news-room/fact-sheets/detail/depression">https://www.who.int/news-room/fact-sheets/detail/depression</a> INTERNET 2%
3	<a href="https://www.mayoclinic.org/diseases-conditions/depression/diagnosis-treatment/drc-20356013">https://www.mayoclinic.org/diseases-conditions/depression/diagnosis-treatment/drc-20356013</a> INTERNET 1%
4	<a href="https://medium.com/@kaanalperucan/mastering-data-preprocessin...">https://medium.com/@kaanalperucan/mastering-data-preprocessin...</a> INTERNET <1%

EXCLUDE CUSTOM MATCHES OFF

EXCLUDE QUOTES OFF

EXCLUDE BIBLIOGRAPHY OFF