KEC_AI_NLP_DEP @ LT-EDI : Detecting Signs of Depression From Social Media Texts

Kogilavani Shanmugavadivel[1], Malliga Subramanian[1], Vasantharan K[1], Prethish GA[1], Sankar S[2], Sabari S[3]
[123]Department of AI, Kongu Engineering College, Perundurai, Erode.
[1]Department of CSE, Kongu Engineering College, Perundurai, Erode.
{kogilavani.sv, mallinishanth72}@gmail.com
{vasantharank.21aid, prethishga.21aid}@kongu.edu
{sankars.21aid, sabaris.21aid}@kongu.edu

Abstract

The goal of this study is to use machine learning approaches to detect depression indications in social media articles. Data gathering, pre-processing, feature extraction, model training, and performance evaluation are all aspects of the research. The collection consists of social media messages classified into three categories: not depressed, somewhat depressed, and severely depressed. The study contributes to the growing field of social media data-driven mental health analysis by stressing the use of feature extraction algorithms for obtaining relevant information from text data. The use of social media communications to detect depression has the potential to increase early intervention and help for people at risk. Several feature extraction approaches, such as TF-IDF, Count Vectorizer, and Hashing Vectorizer, are used to quantitatively represent textual data. These features are used to train and evaluate a wide range of machine learning models, including Logistic Regression, Random Forest, Decision Tree, Gaussian Naive Bayes, and Multinomial Naive Bayes. To assess the performance of the models, metrics such as accuracy, precision, recall, F1 score, and the confusion matrix are utilized. The Random Forest model with Count Vectorizer had the greatest accuracy on the development dataset, coming in at 92.99 percent. And with a macro F1-score of 0.362, we came in 19th position in the shared task. The findings show that machine learning is effective in detecting depression markers in social media articles.

1 Introduction

Millions of individuals throughout the world suffer from depression, a widespread mental health illness that causes personal and social problems. Early detection and response are critical for effective aid and therapy. The rise of social media platforms has offered new options for detecting depression symptoms by monitoring people's online expressions, which might be used for early diagnosis. However, appropriately interpreting these hints from social media postings may be difficult. Due to the massive volume of data and the inherent problems of text analysis, a robust strategy is required. To address this issue, we developed a method in this paper that integrates data pre-processing, feature extraction, and machine learning models to identify depressed symptoms in social media articles. The initial step in our methodology is data preparation, which involves cleaning and preparing the social media posts for analysis. We employ resampling techniques to get over problems with class imbalance and provide a representative dataset. We can lessen biases that may arise from data collection and sampling procedures thanks to this strategy. In the next part, the emphasis changes to feature extraction using popular techniques including Term Frequency-Inverse Document Frequency (TF-IDF), Count Vectorizer, and Hashing Vectorizer. By identifying the text's unique language patterns and word frequencies, these tools allow us to pinpoint relevant qualities for depression diagnosis. To assess the success of our plan, we employ a range of machine learning models, such as Logistic Regression, Random Forest, Decision Tree, Gaussian Naive Bayes, and Multinomial Naive Bayes.

These models are created using the obtained attributes, and they are evaluated for their accuracy in identifying depression or not in social media message classification. According to our findings, the Random Forest model with Count Vectorizer feature extraction had the highest level of accuracy out of all the models we studied. The use of this combination may enable the identification and distinction of language patterns associated with depression, enabling accurate prediction and detection. Everywhere there is an increase in concern over the prevalence of mental health issues, notably depression. Early detection and intervention are

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 300–306, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_046

crucial for effective treatment and support. This gives a potential technique to identify depressed symptoms since more individuals are expressing their thoughts, feelings, and experiences online as a result of the rise of social media platforms. Despite the fact that many tactics have been studied in previous research, there are still several limitations, including the lack of a consistent approach, the significance of context in text interpretation, and the need for automated and scalable solutions. Ethical concerns including privacy and authorization pose questions regarding the use of personal data for mental health detection. It is necessary to create a trustworthy system that can recognize signs of sadness from social media texts while taking contextual factors, privacy difficulties, and ethical considerations into account. A technique like this would help identify people who are at risk early, enabling immediate treatment and assistance to decrease the effects of depression. The pivotal work in (Sampath et al., 2023) not only provided us with valuable guidance to successfully complete the shared task but also empowered us to construct a high-accuracy model.

2 Literature Review

A literature review is a critical and rigorous analysis of academic articles, research papers, and published books that are relevant to a certain subject or area of study. It comprises evaluating, synthesizing, and summarizing prior research and information in order to uncover gaps, contradictions, and trends in the field. A literature review aims to provide a comprehensive account of the existing body of knowledge on a particular topic. It helps researchers find pertinent concepts, theories, and practices, as well as shape their own study design and objectives. It also helps researchers gain a more thorough understanding of the current research environment.

(Hegde et al., 2022) aims to develop automated tools using ensemble machine learning models and transfer learning with BERT to detect signs of depression in social media text. The goal is to improve identification and support for individuals exhibiting depressive behavior. (Victor et al., 2019) discusses about accurately identifying clinical depression using machine learning and automated data collection procedures. The proposed framework combines advanced machine learning techniques with automated data collection to reduce subjective biases and provide a more objective analysis of depression symptoms. In (Islam et al., 2018), the authors discusses about to design a system or model that can accurately identify and classify individuals likely to be experiencing depression based on their social network data. It explores the potential of using machine learning techniques to detect depression symptoms or individuals at risk of depression based on their social network data. In (Liu et al., 2022), the authors suggests directions for future research on using machine learning methods to detect depressive symptoms using text data from social media. Machine learning approaches applied to social media text data can effectively detect depression symptoms, serving as complementary tools in public mental health practice. The research done in (Dinkel et al., 2019) focuses on text-based depression detection in sparse clinical conversations using a multi-task Bidirectional Gated Recurrent Unit (BGRU) network with pre-trained word embeddings. The proposed system models patients' responses during clinical interviews to detect depression severity and binary health state. (Dinkel et al., 2019) addresses about the need for effective depression detection using text-based models and understanding the model's decision-making process. The proposed system is a text-based multitask Bidirectional Long Short-Term Memory (BLSTM) model with pre-trained word embeddings for depression detection and severity prediction. It achieves state-of-the-art performance and provides insights into the words and sentences contributing to predictions. The authors in (Tsugawa et al., 2015) aims to develop an efficient approach using LSTM-based Recurrent Neural Networks (RNN) to identify and predict texts describing self-perceived symptoms of depression. The proposed system utilizes symptom-based feature extraction and outperforms traditional word frequency-based approaches. (Ernala et al., 2019) discusses the lack of reliable and effective emotion detection systems for analyzing and extracting emotions from text data. It surveys approaches, proposals, datasets, strengths, weaknesses, and open issues in text-based emotion detection. The focus is on designing and developing a text-based emotion detection system. (De Choudhury et al., 2014) discuss to develops a metric-based depression detection system using text analysis. It aims to design a metric to describe the level of depression based on text analysis and classify participants accordingly. The proposed system focuses on participant replies for generalized results, but limitations of text-based depression measurement are discussed. The authors in (Guntuku et al., 2017) analyses existing research on detecting depression signs from social media. It focuses on computing tools, linguistic feature extraction methods, statistical analysis techniques, and machine learning algorithms used in the field. The goal is to provide comprehensive information on research papers related to depression sign detection from social media.

3 Methodology

The purpose of this study is to offer a practical method for spotting depressed symptoms in posts from social media. Our dataset is divided into three categories, "not depression," "moderate," and "severe," which represent varying levels of depression severity. Our approach includes feature extraction with TF-IDF, Count Vectorizer, and Hashing Vectorizer as well as the usage of several machine learning models, including Logistic Regression, Random Forest, Decision Tree, Gaussian Naive Bayes, and Multinomial Naive Bayes. We also deal with class disparity by employing resampling methods. The first step in our methodology is data preparation. We clean and prepare the social media messages to make sure they are ready for inspection. This process involves removing unnecessary information, such as URLs, special characters, and numbers. We also employ techniques like lowercasing and stop-word removal to reduce noise and enhance the text data quality. To address the issue of class imbalance in the dataset, we employ resampling techniques. A class imbalance exists when one or more classes are excessively underrepresented in relation to others. Models that are skewed in favor of the dominant class may be the outcome of this discrepancy. To address this, we employ resampling methods like oversampling (like SMOTE) or undersampling (like random undersampling) to balance the classes and give a representative dataset.

Feature extraction, which requires reducing a big collection of characteristics into a smaller, more manageable set, is a crucial stage in machine learning and data analysis. In the field of text analysis, the process of converting textual data into numerical representations that may be used as inputs for machine learning algorithms is referred to as feature extraction. Feature extraction seeks to extract the relevant information from the raw data by eliminating excess or unneeded information. The extraction of key features reduces the complexity of the data, enabling rapid and precise analysis. Feature extraction is essential when unstructured text data is the source for text analysis activities. By utilizing feature extraction techniques, textual data from documents, phrases, or words is converted into numerical representations that algorithms may analyze in text analysis.

Machine learning (ML) models are computer algorithms that extrapolate patterns and predict outcomes from data, as opposed to traditional programming. These models are frequently used in applications such as speech recognition, image recognition, natural language processing, and predictive analytics. ML models have the ability to analyze complex data, identify trends, and make inferences based on the patterns and correlations found in the data. Numerous sectors, including social media analysis, marketing, healthcare, and finance, use ML models extensively. They are able to handle challenging datasets, uncover buried patterns, and provide intelligent predictions and advice. The two primary types of machine learning models are regression models and classification models, each of which focuses on a particular class of problems and has a unique purpose. Regression models are used when the aim variable or result is continuous or numerical.

We evaluate how effective different machine learning algorithms are in identifying depression symptoms. Gaussian Naive Bayes, Multinomial Naive Bayes, Random Forest, Decision Tree, and Logistic Regression are some of the models we employ. These models are trained using the retrieved features and associated class labels. They investigate the best way to divide social media posts into the three categories of "not depression," "moderate," and "severe." We examine the performance of each model using pertinent assessment criteria including accuracy, precision, recall, and F1-score. These metrics reveal how well the models categorize instances into different classes. To make sure that our models are trustworthy and generalizable, we may also employ techniques like cross-validation. After the models have been assessed, we compare their results using the chosen assessment metrics. We identify the model that is most effective at identifying depressive signs. We find that the Random Forest model with Count Vectorizer

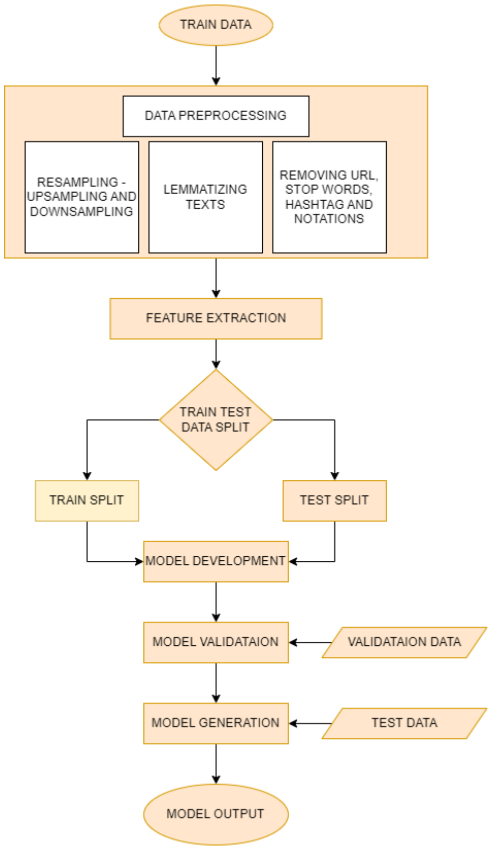

Figure 1: Proposed Model Workflow

feature extraction has the highest accuracy among the models tested in this study. Our research combines data pre-treatment methods, resampling techniques, feature extraction, and machine learning models to effectively identify depression signals from social media articles. The selected Random Forest model offers accurate predictions and helps in the early identification of people who are at risk for getting depression. It employs Count Vectorizer as the feature extraction approach. The general workflow of the system for recognizing indicators of depression is shown in Figure 1.

3.1 Logistic Regression

The logistic regression classification method simulates the relationship between the independent variables and the probability of a certain outcome. It is often used for binary classification problems and may be extended to accommodate multi-class classification tasks. The log-odds of the target variable and the input features are assumed to be linearly connected in logistic regression. The parameters are estimated using maximum likelihood estimation, and the logistic function is used to predict the likelihood of each class. In the context of detecting depressive symptoms based on the characteristics that were gathered from social media texts, the severity of depression may be predicted using logistic regression.

3.2 Gaussian Naive Bayes

The Bayes theorem and the feature independence presumption serve as the foundation for the Naive Bayes classifier, which employs probabilistic classification. A variation of Naive Bayes called Gaussian Naive Bayes assumes that the traits have a Gaussian distribution. The Bayes theorem is used to calculate the posterior probability of each class, and the likelihood of each feature value given the class is then calculated. High-dimensional data can be successfully handled using the computationally effective approach known as Gaussian Naive Bayes. It works well for issues where the independence assumption is partially violated. Gaussian Naive Bayes may be used in the context of this study to categorize social media postings into various degrees of depression severity based on the identified features.

3.3 Random Forest

Many decision trees are used in the Random Forest ensemble learning approach to provide predictions. This versatile and effective strategy may be used to solve classification and regression challenges. Random Forest generates several decision trees by bootstrapping the data and employing random feature groups. The forecasts of all decision trees, each of which was trained on a different sample of the data, are combined to create the final forecast. Random Forest delivers perceptions of feature value, is resistant to overfitting, and excels at processing high-dimensional data. In the context of this study, Random Forest may be used to classify social media messages into different levels of depression severity.

3.4 Decision Tree

Decision Tree, a non-parametric supervised learning system that consists of a hierarchical structure of if-else rules, is trained using data. Each internal node is represented as a feature, each branch as a rule for making decisions, and each leaf node as the result, resulting in a model that resembles a tree. Decision trees can handle both categorical and numerical data and are simple to grasp. To improve the homogeneity of the target variable within each group, they iteratively divided the data based on the most crucial characteristics. Overfitting may occur in decision trees, but it may be prevented by using strategies like pruning, setting a maximum depth, or agreeing on the minimum number of samples per leaf. Decision trees may be used in the context of this study to categorize social media posts into different degrees of depression intensity.

3.5 Multinomial Naive Bayes

Another Naive Bayes variation appropriate for discrete feature variables is multinomial Naive Bayes. It is believed that the features have a multinomial distribution, which is frequently utilized for text classification issues. Using the training data, Multinomial Naive Bayes models each feature value's likelihood given the class, and Bayes' theorem is then applied to get the posterior probability of each class. It is frequently used for text classification tasks including subject classification and sentiment analysis. Multinomial Naive Bayes may be used in the context of this study to categorize social media postings into different degrees of depression severity based on the collected data.

4 Performance Evaluation

The evaluation of model performance is a crucial step in establishing the effectiveness and reliability of machine learning models. To evaluate different aspects of model performance, numerous metrics are utilized. Some examples of regularly employed measures are accuracy, precision, recall, F1 score, and the confusion matrix. These measures are essential for evaluating how well our study's algorithms work at spotting depression symptoms in social media messages. In our work, we trained many models and then used the development dataset to measure their performance. With a score of 92.99%, Random Forest with Count Vectorizer feature extraction was the most accurate model.

This shows that the algorithm correctly predicted the class labels for a large portion of the social media messages in the development dataset. The patterns and characteristics indicative of melancholy in social media communications appear to have been successfully recognized by the Random Forest model with Count Vectorizer feature extraction due to its high accuracy. It is essential to look at additional performance metrics including accuracy, recall, F1 score, and the confusion matrix to get a whole view of the model's performance. By examining accuracy, recall, and F1 score, we can assess the model's capacity to correctly classify texts with indications of melancholy while minimizing false positives. The confusion matrix also provides detailed information on the distribution of the expected and actual class labels, which enables us to identify problem regions and potential misclassification sources. Overall, the development dataset demonstrated the Random Forest model's use of Count Vectorizer feature extraction to achieve the highest accuracy (92.11%). In the tables 1, 2 and 3 below, the accuracy, precision and F1-score are used to compare the performance of various models.

5 Conclusion

To detect signs of sorrow in social media postings, this investigation also used machine learning techniques. The study included the gathering of data, pre-processing, feature extraction, training of the model, and performance assessment. Many models were trained and evaluated using a variety of feature extraction techniques, including Logistic Regression, Random Forest, Decision Tree, Gaus-

Models Used	TF-IDF	Count Vectorizer	Hashing Vectorizer
Logistic Regression	81.93%	86.38%	42.86%
Multinomial Naive Bayes	76.4%	81.82%	41.83%
Random Forest	92.11%	92.99%	91.4%
Gaussian Naive Bayes	74.96%	72.3%	43.68%
Decision Tree	77.57%	79.52%	77.01%

Table 1: Accuracy of Model with Dev Data

Models Used	TF-IDF	Count Vectorizer	Hashing Vectorizer
Logistic Regression	81.9%	86.3%	42.8%
Multinomial Naive Bayes	76.4%	81.8%	41.8%
Random Forest	92.1%	92.9%	91.3%
Gaussian Naive Bayes	74.9%	72.2%	60.8%
Decision Tree	77.5%	79.5%	71.7%

Table 2: Precision of Model with Dev Data

Models Used	TF-IDF	Count Vectorizer	Hashing Vectorizer
Logistic Regression	0.81	0.86	0.41
Multinomial Naive Bayes	0.76	0.81	0.42
Random Forest	0.92	0.93	0.91
Gaussian Naive Bayes	0.74	0.72	0.42
Decision Tree	0.75	0.78	0.74

Table 3: Macro F1-Score of Model with Dev Data

sian Naive Bayes, and Multinomial Naive Bayes. After thorough examination and analysis, the Random Forest model with Count Vectorizer feature extraction achieved the highest accuracy of 92.99% on the development dataset. This shows that the model successfully identified the recurring themes and personality factors linked to depression in social media posts. The study demonstrated how machine learning models can identify depression-related signs in social media data, which can help with early identification and intervention for people who are at risk. Using natural language processing and classification approaches, the models were able to analyse text data and provide insights on the presence and severity of depression symptoms. The results highlight the value of feature extraction techniques like Count Vectorizer in sifting out important data from text input. Additionally, a complete evaluation of the models' effectiveness was provided through the measurement of performance metrics including accuracy, recall, F1 score, and confusion matrix. While the Random Forest model with the Count Vectorizer shown better accuracy, future research should explore novel feature extraction techniques, model architectures, and data sources to enhance the diagnosis of sorrow from social media postings. Overall, this study contributes to the growing body of research on utilizing machine learning for mental health analysis and lays the framework for developing scalable and efficient methods for leveraging social media data to identify and treat depression cases early on.

References

Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638.

Heinrich Dinkel, Mengyue Wu, and Kai Yu. 2019. Text-based depression detection on sparse data. *arXiv preprint arXiv:1904.05154*.

Sindhu Kiranmai Ernala, Michael L Birnbaum, Kristin A Candan, Asra F Rizvi, William A Sterling, John M Kane, and Munmun De Choudhury. 2019. Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–16.

Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.

Asha Hegde, Sharal Coelho, Ahmad Elyas Dashti, and Hosahalli Shashirekha. 2022. Mucs@ text-lt-edi@ acl 2022: Detecting sign of depression from social media text using supervised learning approach. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 312–316.

Md Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M Kamal, Hua Wang, and Anwaar Ulhaq. 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems*, 6:1–12.

Danxia Liu, Xing Lin Feng, Farooq Ahmed, Muhammad Shahid, Jing Guo, et al. 2022. Detecting and measuring depression on social media using a machine learning approach: systematic review. *JMIR Mental Health*, 9(3):e27244.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Jerin Mahibha C, Kogilavani Shanmugavadivel, and Pratik Anil Rahood. 2023. Overview of the second shared task on detecting signs of depression from social media text. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

S. Tsugawa, Y. Kikuchi, F. Kishino, and T. Takahashi. 2015. Monitoring depression via social media: Preliminary results. In *Proceedings of the 2015 IEEE International Conference on Healthcare Informatics*. IEEE.

Ezekiel Victor, Zahra M Aghajan, Amy R Sewart, and Ray Christian. 2019. Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation. *Psychological assessment*, 31(8):1019.