

Study of Early Risks of Depression by Analysing Social Media Posts

IIMS Journal of Management Science
14(1) 9–25, 2023
© The Author(s) 2022
DOI: 10.1177/0976030X221112529
journal.iimshillong.ac.in



Krishna Kumar Singh¹ 

Abstract

Although diagnosing and predicting early signs of depression using social network data is an established position globally, several aspects and dimensions are yet to be detected. This study aims to build an analytical model by considering those new dimensions like the time that can determine whether a user is depressed or not based on the textual data present in the social media posts by the user. The author uses predictive analytical methods to build the model. Classifier models built, detect the sentiment of the text based on the emotion exhibited by the user, which can be categorised as depressive behaviour or non-depressive behaviour based on the sentiment exhibited. The proposed model has displayed an accuracy of 80% and above on the given data. The current proposed model provides the scope to build a system that can aid in reaching out to those who are at the risk of falling into depression or are already suffering from depression for treatment by making them more aware of their emotional state. Practitioners can use it to identify more indicators that can precisely identify symptoms of users exhibiting signs of depression.

Keywords

Emotions, depression, social media, sentiment analysis, predictive analytics

JEL Classification: C11, C53, I11, I12, I53

Received 9 March 2022; revised 9 June 2022; accepted 9 June 2022

¹ Symbiosis Centre for Information Technology, Pune, Maharashtra, India

Corresponding Author:

Krishna Kumar Singh, Symbiosis Centre for Information Technology, Pune, Maharashtra 411057, India.

E-mail: krishnakumar@scit.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

Introduction

Depression is one of the major mental health concerns prevailing worldwide. According to World Health Organization (WHO), more than 264 million people belonging to all demographics globally suffer from depression. Many people rely on social media to connect and share their thoughts in this digital era. Social Media consists of a wide range of websites and applications such as Facebook, Twitter and Instagram. The ever-increasing amount of data available on the Internet has resulted in a massive repository of textual resources. People use social media to describe their experiences and views or simply state everything in their minds. Due to this, social media data is the most valuable source to study the behaviour of human beings for medical purposes. While social media has its benefits, it has also made people feel isolated. Spending excessive time on social media may lead to feelings of frustration, sadness and loneliness, which may affect one's mental health. Heavy usage of social media leads to a high risk of suffering from anxiety and depression, which may lead to self-harming thoughts (Shen et al., 2018). Social media users exhibit behavioural, cognitive and emotional attachment issues with their addictive use of the platform, which leads to tolerance or increased usage with time. Some of the characteristics that contribute to social media impacting one's mental health are fear of missing out, experiencing cyberbullying, developing low self-esteem due to comparison of lifestyles and suffering from sleep disorders. It is important to identify such users directed to a therapist for the right form of treatment. The thought process and behaviour of spending time and sharing data on social media prompted the author to consider social media data for this research. Relationships on social media sites exposed to their close group of friends, colleagues and family, and pressure to project an idealised self-image or avoid presenting a bad image on social networking sites (SNS) may lead to individuals not being honest with their emotional self-disclosure. Thus, social media data that is more reliable in analysing users' emotional states have been collected. It leads to the possibility of identifying users' emotional states based on their posts. Although there are several models for identifying depression from social media data, the model used here is highly trained on textual Data with time as a dimension. Several other machine learning models were also built for this particular data set, but three models with over 80% accuracy have been used for comparative analysis for this research. These machine learning models are robust and require less processing time to generate results. Many times, subjects try to fabricate the data and finding multiple aspects of sentiments are tough from the data of a single source. Data of the same subject available at multiple sources are more authenticated as it reveals more aspects and minimises data fabrication. To consider these issues, the author considered multiple social media sources of data to make the model. In this study, the author examines various linguistic cues which help to detect emotion cause events: the position of cause event and experience relative to the emotion keyword: emotional process such as positive emotion (e.g., 'happy', 'love', 'nice'), negative emotion (e.g., 'worthless', 'loser', 'hurt', 'ugly', 'nasty'), sadness (e.g., 'worry', 'crying', 'grief', 'sad'),

anger (e.g., 'stop', 'shit', 'hate', 'kill', 'annoyed') and anxiety (e.g., 'worried', 'fearful'). A temporal process such as present focus (e.g., 'today', 'is', 'now'), past focus (e.g., 'ago', 'did', 'talked') and future focus (e.g., 'shall', 'may', 'will', 'soon'). Linguistic words such as articles (e.g., 'a', 'an', 'the'), prepositions (e.g., 'for', 'in', 'of', 'to', 'with', 'above'), auxiliary verbs (e.g., 'do', 'have', 'am', 'will'), conjunctions (e.g., 'and', 'but', 'whereas'), personal pronoun (e.g., 'I', 'them', 'her', 'him'), impersonal pronouns (e.g., 'it', 'it's', 'those'), verbs (e.g., 'go', 'good') and negation (e.g., 'deny', 'dishonest', 'no', 'not', 'never') (Al Asad et al., 2019; Islam et al., 2018). There is a great debate over the selection and suitability of available algorithms in machine learning for particular data sets. As the goal of the study is to understand the sentiments of the people and go deep into their mental status, the author selected algorithms that perform better in the case of binary. SVM, logistic regression and Bernoulli naïve Bayes are the most suitable algorithms in this category. Individual use of algorithms has pros or cons, but results are not biased. To make results more unbiased and consider more parameters author considered the results of these three algorithms combined. These algorithms are most suitable for discrete data. Multiple researchers carried out unique algorithms to calculate the sentiment of various subjects, but the robustness with more parameters of those proposed models is still lacking. The proposed model is more robust than the earlier models based upon single algorithms, and the chances of error are relatively less. The research and methodology conducted will be useful in raising awareness among online social network users about their mental health conditions. The observations and approaches will be useful in developing methods for detecting the occurrence of depression for use by healthcare agencies, or on behalf of patients, allowing those suffering from depression to be more informed about their mental health. This study aims to build a machine learning model that classifies the textual data from social media users according to their sentiment through sentiment analysis. This model can be used by researchers who want to explore social media's capabilities in identifying users' mental health conditions, professional therapists or helpline services who want to spread awareness about depression. From a clinical viewpoint, this study offers an immense opportunity to identify users who are at risk of falling into depression or are already suffering from it and then provide appropriate support and assistance to them on the road to recovery by introducing them to legitimate professional help and digital healthcare platforms which offer chatbot services.

Literature Review

In today's SNS generation, more and more people are entering the world of social media to develop connections, both personal and professional. While some users use social media to post their thoughts and emotions, some use it as a medium to promote their business. Either way, one cannot deny that the Data on social media can never be irrelevant and can always be used for several purposes. The future area for the research is how data from social media can be

used to study the relationship between users' posts and psychological constructs and behavioural patterns. Analysing user data for essential information from SNS identifies and emphasises the user's mental health status and condition. As a result of the research, several depressed indicator phrases have emerged, all of which are important in attaining accurate results. The first and greatest priorities for data analysis and diagnosing depressive behaviours were four categories of probable sources that could cause depression, such as emotional state of mind, temporal procedure and word or comment manner from the user's friends (Sonawane et al., 2018). To receive precise and adequate information, more data must be collected. Manual labelling of all complicated attributes via crowdsourcing and deeper dimensions should be examined to construct a better depression detection model (Pandya & Bhattacharyya, 2005). MIDAS is a browser-based analytics platform that examines several aspects of social media users' linguistic and behavioural patterns regarding mental conditions. An online framework that enables the exploration of a user's varied traits about two specific mental conditions is developed to enrich the study further. Some of the approaches have produced limited results that can be utilised to build more complicated structures to understand online user behaviour better. According to the researchers, technology might be used to collect more data from patients and boost the learning model by offering feedback (Vanlalawmpuia & Lalhmingliana, 2020). Data is accumulated through crowdsourcing for users who have been diagnosed with clinical depression based on a standard psychometric test. Users' behavioural cues were used to create a static classifier that provides estimates of the probability of depression before symptoms occur; these distinguishing characteristics were used to create a Classifier model that can predict an individual's possibility of depression before complications arise (Liu & Shi, 2022). The use of social media has allowed for the study of psychological and social concerns on an unprecedented scale. Predictive lexica (words and weights) for age and gender based on word usage in Facebook, blog and Twitter data with relevant demographic labels are produced in this research using regression and classification techniques. Parameters such as age and gender have great importance. at the same time, prediction and lexica are created using a data set of Facebook users who consented to share their status updates and divulge their age and gender (Settanni & Marengo, 2015). To better grasp the underlying reason behind depression, various sorts of KNN classification techniques were being used to detect depression from emotional expression. The research shows that based on linguistic structure, emotional process, temporal methodology and other aspects, these classifiers may effectively extract the depressive emotional outcome (Guntuku et al., 2017). The study provides a taxonomy of data sources and methods that have been used to aid and intervene in mental health. It also focuses on how social networks and other data sets have been utilised to recognise emotions and identify people who might benefit from psychological treatment (AlSagri & Ykhlef, 2020). A depression detection model is built using the characteristics of depressed users collected from clinical examinations. The model captures users' perspectives and sentiment polarity from the texts posted via sentiment analysis. The training data psychologists discovered in this study

is broadly propagated across social media platforms. The study also shows that there is a probability for the connections of depressed users to be depressed, and that various types of interactions generate varying outcomes, indicating that further research is needed (Sap et al., 2014). The author also focuses on the relationship between the textual post shared by the user on Facebook and emotional well-being. In contrast, the studies employing automated extraction methods to textual analytics on the user data have proven feasible compared to the traditional approach. However, the results obtained from such an approach had proved to be mixed and indecisive based on the literature survey performed by the author of the article on different study papers related to exploring the relationship between the linguistics indicators identified from Facebook user-driven posts and emotional well-being. According to the author, this could be due to a lack of consideration of some relevant variables that define user characteristics, which gives light on the user's motive behind social media use. The article has also given an example of how social media use differs based on the users belonging to different age groups. Users tend to post more about their thoughts and use social media to maintain connections. In contrast, older adults use social media for their goals, such as self-promotion, to maintain a certain image of themselves. Due to these existing differences, the article suggests the researchers include age as an intervening variable when analysing the SNS data (Guntuku et al., 2017). Although depression has an associated high treatment rate, approximately two-thirds of people who suffer from it just do not constantly pursue or undergo medical care. Individuals will become more aware that they will need to seek assistance if there is a more comprehensive method for detecting depression and foresee associated risks. A machine learning model is used in this study to assess levels of depression by examining users' posts on social media. The model has been presented via Twitter and Facebook postings. There are several variables to consider while assessing whether or not a person is depressed. To showcase the performance of the models, the findings obtained have been compared to an online question-and-answer survey with participants. Another significant difference is that all posts have been accumulated irrespective of keywords and phrases specifically targeting an individual. Consequently, the model used could more accurately identify depression in any individual (Sadeque et al., 2018). The use of SNS and emotional resonance could impact sleep quality and depression degrees. The degree and severity of depression have been classified as an outcome of this study. This article includes a training data set, and social media posts are subdivided based on the training set. These could also transmit an alert or notification to the user's family (Calvo et al., 2017). Most current studies employ a singular approach to detection, and the class imbalance distributions frequently result in lower detection accuracy. Moreover, in high-dimensional data sets, a significant number of noisy and irrelevant variables hamper the detection accuracy rate. Hybrid feature selection and stacking ensemble techniques are also useful for identifying depressive individuals to deal with uncertainties. To determine feature importance, mutual information value, construct feature weight vector and the iterative elimination technique and extremely randomised

trees approach has been utilised (Khilwani et al., 2021). Because depressive individuals tend to be younger, most users in the microblogging depression networks turn out to be teenagers, and the data set examined may not have been indicative of the entire population. Those in the control group, on the other hand, could be suffering from other psychological conditions. This study provides a preliminary assessment of certain individuals who have been having problems distinguishing whether or not they are unwell (Kumar et al., 2022). The problem of enhancing identification is particularly targeted to Twitter data as the source domain has been explored. It starts by systematically analysing depression-related features across subject areas and summing up significant identification obstacles, isomerism and divergence. A cross-domain Deep Learning-based, Neural Network model with Feature Adaptive Transformation & Combination (DNN-FATC) strategy for transferring key data across heterogeneous domains has been proposed (DBSA, n.d.). In the Annals of research conducted, it can be inferred that it is difficult to accurately comprehend various symptoms and circumstances of individuals through SNS. By integrating data from social media with clinical data, researchers aim to improve detection accuracy even more. It is challenging to undertake large-scale deep learning algorithm development due to depressed individuals' sample data insufficiency. Depression-like symptoms also increase during COVID-19, which has also been reflected in social media posts (Rathee & Kumari, 2022). Researchers can extend the data set's specialisation and diversity in future studies as an outcome. In terms of data may provide customised research for various user types, such as gender, age and family structure.

Methodology

The steps mentioned in the framework, in Figure 1, have been followed to carry out data analysis for the data collected from the SNS (Twitter, Facebook and flicker) and build a model to classify posts to detect tweets indicating signs of depression.

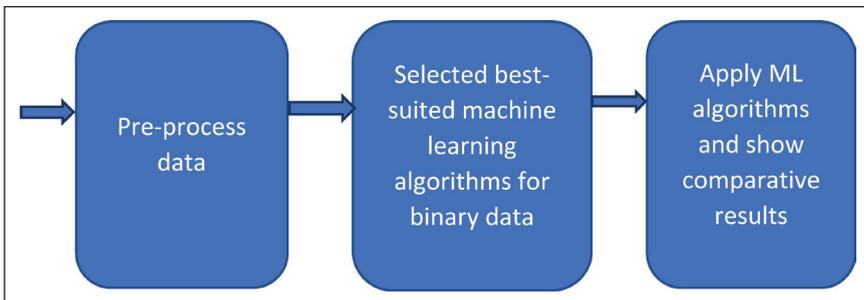


Figure 1. Conceptual Framework.

The social media (Twitter, Facebook and flicker) data has been used in this study through respective APIs (Twinesocial, share count and hashtags). The author collected data from multiple sources during the same period. If a person is suffering from depression or symptoms, that have been generated, he has a habit of sharing similar emotions on different channels. So, the period is paramount to collect data to understand emotions with greater depth. Posts have a variety of distinguishing characteristics, for example, Twitter posts are restricted to 140 characters in length, and Facebook posts have no restrictions. A summary of data with characteristics and quantity is given in Table 1. It is easy and convenient to gather millions of posts for building a model by using the social media site API. Netizens post on social media from various sources, including their mobile devices such as android or iPhones. Spelling errors and slang were far more prevalent in posts obtained from other domains. The quantity of posts does not mean several people but several posts. Sometimes, people post multiple times on multiple sites simultaneously. In this article, the author uses social media APIs to collect data and test models.

For the analysis of data, categories of data are required. Table 2 is the breakdown of target terms into various categories.

The post has been marked with the values based on the sentiment they exhibit through Sentiment Analysis. It is a method of identifying if a text contains neutral, negative or positive emotions. It is one of the subsets of textual analytics that uses both Natural Language Processing (NLP) and machine learning. Only English-language posts were included in the data set. Post-processing has been done on the data set. Emojis have been eliminated for training. Positive and negative emoji filters have been filtered out and omitted from all posts. Reposts

Table 1. Summary of Data.

Characteristics	Quantity
Total number of cells	180,062
Number of depressive indicator cells	92,567
Verbs	6,027
Prepositions	4,928
Interrogative	2,563
Temporal format	DDMMYYYY:HHMMSS

Table 2. Summary of Social Media Text Data.

LIWC Derived Cues	Example Word
Positive words	Happy, love, lovely, sweet
Negative words	Worthless, loser, hurt, ugly, 'nasty'
Sorrow words	Worry, crying, grief, sad
Anger words	Stop, shit, hate, kill, annoyed
Anxiety words	Worried, fearful

and quoted posts that were duplicated from another account have been deleted. Post exhibits of both positive and negative emotions have been collected to make the model non-biased.

The data set contains six columns, namely

1. Target: It indicates the polarity of the posts collected, where 0 indicates that the tweet has negative polarity, and 4 indicates positive polarity.
2. Ids: It indicates the id of the posts.
3. Date: It indicates the date and the time on which the posts have been posted.
4. Flag: The value is NO_QUERY if it does not contain any query.
5. User: It contains the user ids of the accounts concerning the posts gathered.
6. TweetText: It contains the actual text post that has been posted.

Before moving to Data Pre-processing, columns that are not required for the model building have been dropped, and only two columns, that is, target and post text, have been used to carry out further Data Pre-processing. In this stage, the data has been analysed for missing values first. After confirming that data has no missing values, data cleaning has been carried out by removing noisy data from the tweets. This has been followed by text tokenisation and normalisation. These are NLP techniques. Text tokenisation is a method in which the text is divided into tokens, and normalisation is a process in which a word is transformed into its basic form. During pre-processing following steps have been carried out to handle noisy data to avoid errors during model building:

1. Converting text to lower case.
2. Replacing the links such as www, HTTP or HTTPS present in the textual Data with URL.
3. They were handling data containing emojis, which are often used by netizens to display their emotions by replacing them with a pre-defined lexicon that contains the meanings of the emojis used.
4. Non-Alphabets have been eliminated by substituting space for all characters, excluding Digits and Alphabets.
5. Consecutive characters have been removed and replaced with two letters if there are up to three or four characters.
6. Short Words that are words with a length of fewer than two have been eliminated.
7. Stopwords used in the English language that do not often add much meaning to the sentence have been eliminated.
8. Lemmatising has been carried out. Transforming a word to its base form is known as lemmatising.

Feature Extraction

To describe and compare depressive and non-depressive posts, the author extracts the different features given psycholinguistic measurements from the user's post. It is clarified briefly as follows:

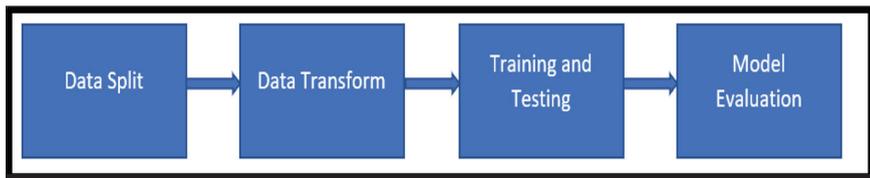


Figure 5. Data Model.

Data has been visualised to find out the most common words in tweets containing negative sentiment value, that is, tweets indicating signs of depression.

From the word cloud visualisation in Figure 4, it can be seen that tweets indicating negative sentiment have words indicating stress, and insomnia frequently appearing, such as miss, sleep, work and sick.

Model Building

The following steps have been carried out during the model building process (Figure 5).

Training and Testing of Data for Machine Learning

The data need to be split into train and test data for an unbiased assessment of the efficiency of the prediction model. The data set has been divided into test and train data by carrying out splitting in which training data accounted for 70% and test data for 30%.

```
X_train, X_test, y_train, y_test = train_test_split (precessestext, sentiment,  
test_size = 0.30, random_state = 0  
Print (data Split done.)
```

Followed by which TfidfVectorizer has been instantiated. It is an effective method for simplifying text documents by converting text in tweets into numerical arrays. It aids in handling the most common words. The word counts are evaluated by measuring how frequently they appear in the documents.

```
X_train = vectorizer.transform (X_train)  
X_test = vectorizer.transform(X_test)  
Print (f'Data Transformed.)
```

For transforming the data set, TF-IDF Vectorizer has been utilised, and X train and X test data sets have been transformed into a matrix of TF-IDF Features. These converted data sets have been used to train and test the models.

Model Evaluation

The data set used for the study is not skewed. The evaluation metric chosen for the model is accurate. It is a key performance indicator that distinguishes a good classification model from a bad one, especially when evaluating a binary classifier. It refers to the total percentage of occurrences that have been successfully predicted.

It consists of four essential components:

1. TP (True Positives): This represents the total number of occurrences in the positive class that has been successfully predicted.
2. TN (True Negatives): This represents the total number of occurrences in the negative class that has been successfully predicted.
3. FP (False Positives): This is the total number of occurrences that have been anticipated to belong to the positive class but ended up in the negative class instead. It is also known as type 1 error.
4. FN (False Negatives): This is the total number of occurrences that have been projected to be in the negative class but are instead in the positive class. It has also known as type 2 error.

$$\text{Formula: Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

BernoulliNB Model

Based on the Bayes' Theorem, Naive Bayes classifier is a set of classification algorithms. The algorithms follow a common idea that each pair of features being classified is independent of the others. The conditional probability of an event A given event B is determined using the Naive Bayes Classifier. BernoulliNB executes Naive Bayes training and classification algorithms for data distributed according to multivariate Bernoulli distributions. There may be multiple parameters, but each is considered a binary-valued (Bernoulli) variable.

```
BNNmodel = BernoulliNB(alpha = 2)
BNBmodel.fir(X_train, y_train)
Model_evaluate(BNBmodel)
```

The accuracy obtained using the Naïve Bayes Classifier model is 80% approximately. Results of bernoulliNB model on the dataset is given in Table 4.

Table 4. Bernoulli Naive Bayes Model.

	Precision	Recall	F1_score	Support
0	0.81	0.79	0.79	39989
1	0.80	0.81	0.79	40011
Accuracy			0.79	80000
Macro avg	0.80	0.80	0.79	80000
Weighted avg	0.80	0.80	0.79	80000

Table 5. Logistic Regression Classifier Model.

	Precision	Recall	F1_score	Support
0	0.83	0.82	0.83	39989
1	0.82	0.84	0.83	40011
Accuracy			0.83	80000
Macro avg	0.83	0.83	0.83	80000
Weighted avg	0.83	0.83	0.83	80000

Logistic Regression Classifier

It is widely utilised for Binary Classification problems. It is based on the logistic function, also commonly known as the sigmoid function. It inputs any real-valued number and projects it to a value between 0 and 1. For a given collection of features (inputs), the target variable (output) can only take discrete values in the case of classification.

```
LRmodel = LogisticRegression(C = 2, max_iter = 1000, n_jobs = -1)
LRmodel.fit(X_train, y_train)
Model_evaluate(LRmodel)
```

The accuracy obtained using the Logistic Regression Classifier model is 83% approximately. Results of logistic Regression model is given in Table 5.

Linear SVC Model

The Linear Support Vector Classifier (SVC) approach uses a linear kernel function to classify data. This approach is feasible for a data set with a huge number of samples. In contrast to the SVC model, Linear SVC consists of more features, such as penalty normalisation ('L1' or 'L2') and loss function.

```
SVCmodel = LinearSVC()
SVCmodel.fit(X_train, y_train)
Model_evaluate(SVCmodel)
```

The accuracy obtained using the Linear SVC model is 82% approximately.

Built models have been stored using the PICKLE function. It saves the serialised machine learning models to a file. It can be deserialised to make further or new predictions. Results of linear support vector machine model on the dataset is given in Table 6.

Interpretation of Analysis & Results

As seen from result in Figure 6, compared to the other two models', the logistic regression model turned out to be the most accurate model for the given data set. While the other two models were close in terms of accuracy but when it comes to runtime, BernoulliNB is the model that was able to perform the classification fastest with an accuracy close to the other two models.

Table 6. Linear Support Vector Machine Model.

	Precision	Recall	F1_Score	Support
0	0.82	0.81	0.82	39989
1	0.81	0.83	0.82	40011
Accuracy			0.82	80000
Macro avg	0.82	0.82	0.82	80000
Weighted avg	0.82	0.82	0.82	80000

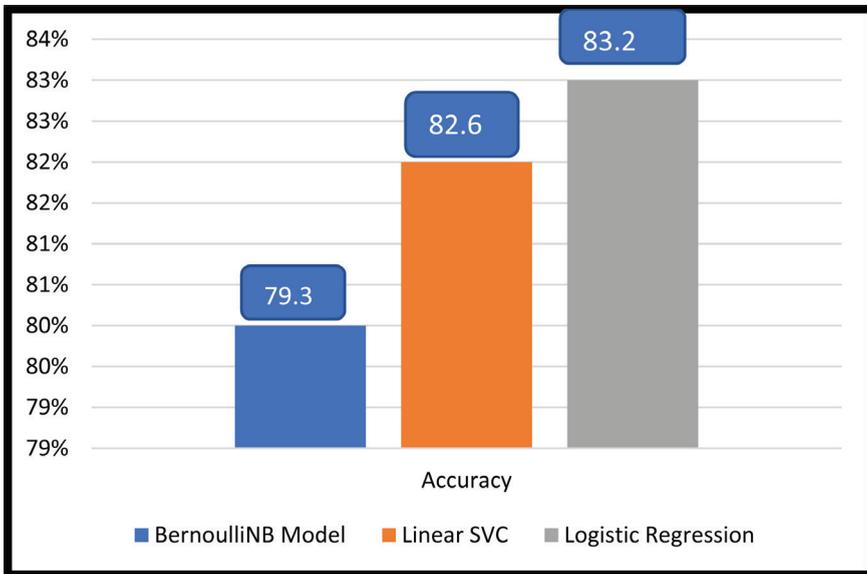


Figure 6. Comparison of Accuracies of Three Machine Learning.

Table 7. Sentiment Model for Identifying Sentiment of the Textual Data.

	Text Sentiment
0	I hate life
1	May the world be a better place
2	I feel empty

Text = ['I hate life', 'May the world be a better place', 'I feel empty'].

The vectorizer has been used to convert data into a TF-IDF Features matrix. This model can be used to detect the emotional state of the textual tweets or any other data in textual form. As seen from the results in Table 7, the model identifies the sentiment of the text and classifies it accordingly.

Conclusion

Mood swings, withdrawal symptoms, antagonism and relapse are all common addiction signs in those who use SNS. It is usual for one's emotional state to fluctuate due to their use of SNS. Depression is a complex subject and several times has been misdiagnosed as it has often been mistaken with symptoms such as regular anxiety and stress. While all these factors combined could lead a person to fall into depression but to be able to diagnose it accurately using data models is still a vast area that needs to be further studied by considering various other factors by examining the mental health of users active on social media and also users who are rarely active on social media. This study displays how textual data from social media can be a valuable medium for identifying signs of depression by building a predictive model from the social media data set. Three models have been built: BernoulliNB Model, Linear SVC Model and Logistic Regression classifier model. The models have been trained based on the sentiment value with the time factor of profile collected from social media. The proposed model successfully identifies users suffering from early signs of depression which can be used to direct them to a therapist for the correct form of treatment. Data of the same subject available at multiple sources are more authenticated as it reveals more aspects and minimises data fabrication. The author took data from multiple social media sites to make the model more robust. The author has considered more parameters for these three algorithms to make the results more unbiased. The findings from the model can be further used to explore the user whose tweets exhibit negative sentiment for identifying potential risks of falling into depression. The author also observed that subjects often tried to show symptoms slowly or intentionally share data on a single platform to reap other hidden benefits. To negate these discrepancies, the author focused more on the data shared on multiple platforms during the same time period. The model can also negate intentionally manipulated value posts on social media while considering textual data. The prediction results of these models are encouraging and helpful in predicting early signs of depression. Symptoms of early depression growing into the subject are not visible, and this model helps detect symptoms in the very early stages. All classification models resulted in an accuracy of over 80%. For scientific improvement, the model can be tested to boost the accuracy by increasing the training data, which can be done through social media analytics by scrapping for more textual data from social media users. This resulted in better results in diagnosing symptoms for medical aspects. For future developments, through this research, models can be further developed by considering other indicators such as the temporal aspects at which a post has been made and other indicators that would give more insight into the emotional state of a user making posts on social media.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Krishna Kumar Singh  <https://orcid.org/0000-0003-3849-5945>

References

- Al Asad, N., Pranto, M. A. M., Afreen, S., & Islam, M. M. (2019, November). Depression detection by analyzing social media posts of user. In *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)* (pp. 13–17). IEEE. <https://doi.org/10.1109/SPICSCON48833.2019.9065101>
- AlSagri, H. S., & Ykhlef, M. (2020). *Machine learning-based approach for depression detection in Twitter using content and activity features*. *IEICE Transactions on Information and Systems*, E103.D(8), 1825–1832. <https://doi.org/10.1587/transinf.2020EDP7023>
- Calvo, R., Milne, D., Hussain, M., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, 23(5), 649–685. <https://doi.org/10.1017/S1351324916000383>
- DBSA. (n.d.). *Media: DBSA Facts about Depression*. [secure2.convio.net](https://secure2.convio.net/dabsa/site/SPageServer/?jsessionid=00000000.app20101a?NONCE_TOKEN=B2DB3A070406934661E7561D). https://secure2.convio.net/dabsa/site/SPageServer/?jsessionid=00000000.app20101a?NONCE_TOKEN=B2DB3A070406934661E7561D
- Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, 18, 43–49. <https://doi.org/10.1016/j.cobeha.2017.07.005>
- Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*. <https://doi.org/10.1007/s13755-018-0046-0>
- Khilwani, V. O., Gondaliya, V., Patel, S., Hemnani, J., Gandhi, B., & Bharti, S. (2021). *Diabetes prediction, using stacking classifier*. 1–6. <https://doi.org/10.1109/AIMV53313.2021.9670920>
- Kumar, P., Samanta, P., Dutta, S., Chatterjee, M., & Sarkar, D. (2022). Feature based depression detection from Twitter data using machine learning techniques. *Journal of Scientific Research*, 66(2), 220–228.
- Liu, J., & Shi, M. (2022). A hybrid feature selection and ensemble approach to identify depressed users in online social media. *Frontiers in Psychology*, 12, 802821. <https://doi.org/10.3389/fpsyg.2021.802821>
- Pandya, A., & Bhattacharyya, P. (2005). Text similarity measurement using concept representation of texts. *International Conference on Pattern Recognition and Machine Intelligence*. Springer.
- Rathee, V., & Kumari, S. (2022). impact of virtual try-on technology on customer's mental imagery during COVID-19. *IIMS Journal of Management Science*, 13(1). <https://doi.org/10.1177/0976030X211051095>

- Sadeque, F., Xu, D., & Bethard, S. (2018, February). Measuring the latency of depression detection in social media [Paper presentation]. *The Eleventh ACM International Conference on Web Search and Data Mining* (pp. 495–503). ACM.
- Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., Ungar, L., & Schwartz, H. A. (2014). Developing age and gender predictive lexica over social media [Paper presentation]. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1146–1151. <https://doi.org/10.3115/v1/d14-1121>
- Settanni, M., & Marengo, D. (2015). Sharing feelings online: Studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology, 1045*. <https://doi.org/10.3389/fpsyg.2015.01045>
- Shen, T., Jia, J., Shen, G., Feng, F., He, X., Luan, H., Tang, J., Tiropanis, T., Chua, T. S., & Hall, W. (2018). *Cross-domain depression detection via harvesting social media* (pp. 1611–1617). <https://www.ijcai.org/proceedings/2018/223>
- Sonawane, N., Padmane, M., Suralkar, V., Wable, S., & Date, P. (2018). Predicting depression level using social media posts. *The International Journal of Innovative Research in Science, Engineering and Technology, 7(5)*, 6016–6019.
- Vanlalawmpuia, R., & Lalhmingliana, M. (2020). Prediction of Depression in Social Network Sites Using Data Mining [Paper presentation]. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 489–495. IEEE.