ISSN: 0974-8571 Vol.14 No. 1 June, 2022

International Journal of Computational Intelligence in Control

A Hybrid Method for Stress and Depression Detection from Social Media Text Focused on Local Languages

Nasir Gul[1], Aurangzeb Khan[1,2], Mazliham Mohd Su'ud[2,3], Muhammad Mansoor Alam [2,3,4], Khair Ullah Khan[1]

[1]Department of Computer Science, University of Science and Technology, Bannu Pakistan

[2]Faculty of Computer and Information, Multimedia University, Cyberjaya, Malaysia

[3]Malaysian Institute of Information Technology University Kuala Lumpur, Malaysia

[4]Faculty of Computing, Riphah International University, Islamabad, Pakistan

Abstract

Sentiment analysis has been increasing enormously in recent times due to the broad range of business and social applications. Sentiment analysis from unstructured text has recently received significant attention from the research community. Now a days people like to comment and share their views on social networks. The diversity of social network is an ideal way to extract the sentiments of the people. People use social media to express their emotions and comment on most of the events occurred around them. This work will focus on determining emotion in line with context dependence within the text. The main challenge is the detection of multilingual text in sentences in the domain of stress and depression focused on the local languages. In this study various machine learning and deep learning model are tested on benchmarks datasets. The machine learning model tested includes support vector machines, random forest and the recurrent neural network. The average accuracy the sentiment model provides is nearly about 80%.

1. Introduction

Depression is classified as a mental disorder. The world health organization has declared depression as the world's 4th major cause of disability (Meitei and Singh, 2019). Furthermore, there exists a significant positive relationship between stress and depression, with stress being a potential cause of depression (Chiba et al., 2012). Despite the psychological effects, as Stanton et al. (2020) indicate, depression also adversely affects the physical health of the respective individual. For example, according to Lee et al. (2015), stress and depression can be a potential cause of a poor and weak immune system which eventually leads to a vast number of critical and life-threatening diseases, with gastric cancer being one of the widely researched.

In today's highly competitive society, stress is a huge issue. People might be stressed for a variety of reasons, including social, economic, political, or personal reasons such as losing a loved one or the fear of being ignored by others. If managed well, stress can keep one focused and inspired to accomplish an excellent job. However, if a person is unable to handle stress appropriately, it can progress to depression, which can cause long-term mood changes plus bad physical conditions. (Hammen, 2015)

Keeping in view the severe and critical consequences of stress and depression, the early detection of this disorder via the relevant and associated prevalent symptoms holds critical importance. Early identification of depression plays a vital role in an appropriate treatment to reduce the adverse effects associated with both physical and mental health. However, as Cacheda et al. (2019) indicate, there exist limited and a little number of services and tests that provide an accurate prediction and detection of a prevailing critical mental health disorder.

With the advancement of Machine Learning and Data Analytics tools, it may now be possible to detect or even forecast stress and depression. Social media can be a strong tool in this regard, as its use has skyrocketed in recent years, regardless of age, occupation, or gender. Social media networks such as Twitter, Facebook, and Instagram have become a significant aspect of almost everyone's life in today's global world. People engage with these networks primarily in terms of sharing their points of view, feelings, thoughts, and emotions. The exponential amount of data being generated via these social media platforms has made these a potential research topic, particularly in terms of text mining. As indicated in the Study conducted by Gaind, Syal, and Padgalwar (2019), text mining of social media content can be effectively be utilized in determining the emotional state of an individual. This technique is proven to be very beneficial, is widely utilized by marketers to understand the behavior of their target audience and design effective strategies accordingly (Seyeditabari, Tabari, and Zadrozny, 2018).

User generated content (UGC), i.e. users' status, tweets, photographs, videos, and social activity data are often analyzed using an automated depression diagnosis algorithm based on machine learning. The classification findings of the predicted objects are then displayed, with the majority of them being presented as a binary result of regular or depressive. Thus, if a person is suspected of having a depressed disposition, additional resources and treatment, including later medical and psychiatric diagnosis, can be provided earliest. Because they can handle a vast quantity of instant interactive user data, heuristic learning algorithms like these are pretty successful for the early diagnosis of depression.

Contribution of the Research: The main contribution of this work are:

- The development of an enhanced lexical pattern-based algorithm for the extraction of features from Emotion text domain. This process is based on linguist patterns linked with emotions words. For the said purpose, a pragmatic examination will be performed to discover how the precision of features extraction is related to the lexical patterns.
- Development of Identification SD Identification Matrix to identify potential stress and detection emotions texts to see how they are represented in the local languages.

The rest of the paper is organized in seven sections such that Related work is provided in section 2, proposed model is presented in section 3, section 4 describe results and discussion, recommendations are provided in section 5, future direction is provided in section 6 while section 7 conclude the findings.

2. Related Work

The conventional method which was used to detect the depression is referring the patients to the clinical depression criteria. With the prevalence of social media networks people got more inclined towards sharing their emotions on social media. Thus, the harvesting the social media is commonly used to detect those expressions and classify it. Therefore, this part is significant as it comprises of the major reasons which led to the need of detection of the depression with an addition of the key concept which lies behind the machine learning techniques that has been extensively used for this purpose. This part also comprises of integration of machine learning which is widely used for various purposes and hence gain its importance in detection as well. Further different techniques which have been used for detection is also studied and included in this chapter. Moreover, this chapter is associated with critically analyzing the reliable published papers in order to support the content of the respective topic. The researcher engages in a thorough discussion regarding the challenges associated with the detection of stress and depression by utilizing machine learning techniques. The use of four classification algorithms Support Vector Machine, Random Forest, Linear Regression and Naïve Bayes is used. The accuracy of data are further enhanced by applying 10-Fold Cross-Validation. (Shatte, Hutchinson et al. , Shatte, Hutchinson et al. 2019) discussed machine learning (ML) applications for mental health, highlighting most recent research and applications in practice. (Sau and Bhakta 2019) uses five machine learning classifiers i.e., Logistic Regression, CatBoost, , Naïve

Bayes, Random Forest, and Support Vector Machine, were evaluated using the Python programming language and the results were promising. A model was introduced for voting systems which is trained using three classifiers (multinomial naïve bayes, gradient boosting, and random forest) and popular voting using an ensemble voting classifier is done. Preliminary results are evaluated for tweet(Kumar, Sharma et al. 2019). (Rajput and Ahmed 2019) discussed the role of social media platform to gain valuable insights into user behavior. They used representative corpus for NLP and also discussed the challenges face today that exists on the internet. Their focus was on building a corpus from social media that is focused on detecting mental illness. They also presented a case study and demonstrate the effectiveness of using such a corpus for helping practitioners detect such cases.

2.1 The Existing Sentiment Analysis Systems

Sentiment Analysis methods and approaches are thoroughly investigated and compared. The comparisons were made on the basis of the used datasets in these systems, the research algorithms/methods that are used for conducting sentiment analysis and opinion mining, application areas and many other features. The comparisons among various latest sentiment analysis system have been made in this section in represented in table 2.

Table 2 represents the assessment of various sentiment analysis systems. In this table, the 1st column holds the serial number of various systems, 2nd column contains the corresponding references of each sentiment analysis system, 3rd column shows the algorithm(s)/methodology that have been used for these systems, 4th column is reserved to the datasets used for sentiment analysis, section 5 contains the results in terms of various assessment metrics but many researchers didn't mentioned their evaluation results,

S. No	Reference	Methodology/ Algorithm/	Dataset used	Results
1.	L. P. Spear, (2000)	None	None	None
2.	yuan Xue, Qi Li , Li Jin, Ling Feng, David A. Clifton,Gari D. Clifford, (2013)	Single tweet based pressure detection, NB, SVM, ANN	10000 tweets	None
3.	Munmun De Choudhury ,Scott Counts, Eric Horvitz (2013)	tHelper, Gaussain Process	Twitter	
4.	Steve chancellor (2019)	CSC	-	None
5.	Moin Nadeem, (2016)	Major Depressive Disorder (MDD) Adaptor	Twitter	Accuracy 84%
6.	Sharath Chandra Guntuku (2019)	EasyAdpt, ANN	Twitter	Accuracy 82%
7.	Fatima, Li	DASentimental, Depression	142 suicide notes	Accuracy 85%

	,Stella (2021)	Anxiety Stress Scale		
8.	A. Kumar (2019)	social media verbalization	1000 Tweets	Accuracy 95%
9.	Li, Liu (2020)	decision tree, random forest, AdaBoost, LDA, and K-nearest neighbor	1500 Tweets	Accuracy 98.38%
10.	Richter, Fishbain ,Markus,Levin & Hadas Okon-Singer B. (2020)	C. Standard Power Analyses	Sample Size of 125 participants	None
11.	Shumaila, Noor ul Huda, Amin, Samina Khalid D. (2022)	E. Classification, Deep Learning and Ensemble.	637 patients	
12.	Patel,Khalaf, Aizenstein, H.J (2016)	SVM, RVR	Twitter	Accuracy 80%
13.	Su, Zhang, He, K.Chen, (2021)	LR, SVM, DNN with LSTM	1539 patients	Accuracy 95.4%

Semantic analysis is a much admired field with a huge number of articles published every year. In fact, sentiment analysis may be categorized into four main areas that are as follows:

1. The sentiment analysis is not only limited to text analysis but it varies up to emotion detection and sentiment strength detection.
2. Sentiment analysis can be conducted at different levels of granularity, that are word-level, document-level, sentence-level, etc.
3. Sentiment analysis already using diverse approaches such as lexicon based sentiment analysis, machine learning algorithm based on these two approaches.
4. The data for opinion mining mainly extracted from social media networks (like Facebook. tweeter etc.) and also from text (e.g., product reviews and news articles etc.).

3. Proposed Model

Generally, data analysis and classification is recognized as a dual-step technique in which the initiative is the training phase, where a classification model develops the classifier by seeking from a dataset composed of databank tuples and are linked with labels (Dimitris Effrosynidis). A label is nothing but a distinct value and each and every value serves as a class (Duyu and Furu, 2014). The second and the most important step is to manipulate the model for classification and then the accuracy and efficiency of the classifier is predicted and estimated (U.A Siddiqua et al., 2016; Gautam and Yadav, 2014).

Nowadays the most widely technique used for sentiment analysis is ML (Machine Learning) and AI (Artificial Intelligence) and in this study, the main focus is on how sentiment analysis explores its possible methods in Roman Urdu. Mood detection by using only words. The average accuracy the sentiment model provides is nearly about 80%. There are multiple algorithms that are for mood prediction like SVM and Random Forest. Most of the papers mentioning the tweeter, Facebook and Instagram data has been used for sentiment analysis and the big tech company's used this big data and analyze this data to check the mood and emotions of the user towards a particular thing. To check the mood and opinions of people in 2020 from January to May on 'Metoo' hashtags and also from the women hashtags the

Metoo there are 50429 tweets and for hashtags Women, there are about 50430 and most of the tweets wherein January and the lowest number in a tweet for both is on February (Kaur, 2021). As mentioned above the Python programming language best fits for data analysis and categorization so, for Sentimental Analysis, data can be classified and analyzed using different Machine learning algorithms and techniques. The method for given sentimental analysis is the Random Forest Classifier of Machine learning (U.A Siddiqua et al., 2016; Gautam and Yadav, 2014).

3.1 System Flow

A ton of corpora are produced for Emotion Detection assignments for Arabic, English, French and other European dialects. Nonetheless, for South Asian dialects particularly Urdu, there is nonexistence of typical corpora accessible. In this article, we present a dataset called RUED, which is openly accessible for the utilization of scholarly exploration reason as it were. Our Roman Scripted Emotion corpus includes 10,000 sentences. Corpus is physically hand-marked. The source sentences are gathered from various assets. Our assets can be found through $Github^3$. [1]

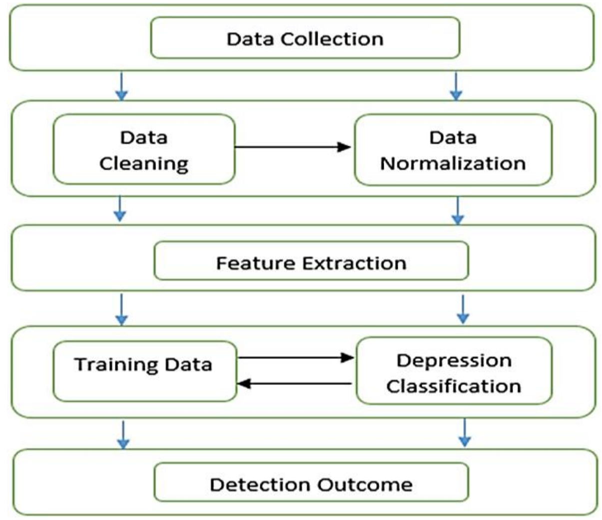

Figure: System flow Diagram

A lot of corpuses presented and developed for Emotion Detection tasks for Arabic. English, French and other languages. However, for most of South Asian languages like Urdu, there is a lack of standard corpora availability. In this article, a new dataset called RUED which is a roman Urdu Emotion Detection corpora, which will be available for the use of research Community. To the best of our knowledge, RUED is Roman Urdu's first presented dataset for emotion detection analysis.

3.2 Data Collection

3.2.1 Data Pre-Processing:

Algorithm for Normalizing Roman Urdu

1. Scrap data from social media websites into a text file
2. Clean the existing raw data
   - a. Removal of Noise.
   - b. Removal of hash tags, Retweets and URLs.
   - c. Removal of additional spaces preceding a line of text
   - d. Removal of additional spaces following any line of text
   - e. Removal of non English characters and numeric values
   - f. Replace multiple spaces by a single space.
   - g. Consistency in Spelling
   - h. Remove punctuation marks
   - i. All words converted in Lower-case
3. Read input from text file.
4. Compare each and every string with a list of Proper Nouns [2]

3.3 Data Annotation:

The text annotation is based on 4 emotions namely, Happy as h, Angry as a, Sad as s, Neutral as n. The annotated dataset is taken from [1]

Assigned an emotion to one example from the above described four emotions list.

- If one example does not belong to any above class of emotion, then assigned neutral class to that example.
- If an example belongs to multi-category of emotions, then they assigned an emotion which is particularly closer to that example.
- Allocated one feeling to one model from the above described four feelings list.

- If one model doesn't have a place with any above classification of emotion, then, at that point they appointed impartial class to that example.
- If a model has a place with multi-classification of feelings, then they allocate to a feeling which is very nearer to that model.

3.4 Characteristics of Corpus:

After the consummation of explanation measure, the conduction of statistical examination on corpus to distinguish any qualities of emotional articulations in Roman Urdu. The essential insights of the corpus is displayed in table II a,b,c. It is explained 10,000 sentences out of 20,000 figurative sentences including feelings. There are feelings in a significant level of Roman Urdu's allegorical sentences in the corpus. Figure 4 exhibits the feeling classification dissemination in

The corpus. In Table III, a few instances of RUED shown in which we introduced 3 instances of every feeling class.

Table 2 a. Whole Corpus Statistics

Corpus Statistics	
Count of Sentences	20000
Count of Total words	360000
Unique counts	30000
Average number of words in sentences	18

Table 2 b. Annotated Corpus Statistics

Annotated Corpus Statistics	
Count of Sentences	10000
Count of Total words	150000
Unique counts	17590
Average number of words in sentences	14

Table 2 c. Emotion Corpus Statistics

Emotion Corpus Statistics	
Total number of emotions	4
Count of Happy sentences	3000
Count of sad sentences	2300
Count of angry sentences	1700
Count of neutral sentences	3000

3.5 Feature Extraction:

To portray & exhibit among depressive and non-depressive posts, firstly we remove the various elements considering psycholinguistic estimations from the client's post. It is explained momentarily as follows:

Psycholinguistic provisions LIWC is a psycholinguistic vocabulary bundle made by mental examiners to see the distinctive full of feeling, scholarly, and etymological parts lies on client's verbal or composed correspondence

3.6 Linguistic process:

Linguistics process is one of the major parts of LIWC psycholinguistics vocabulary package. It was proposed to measure word use in intellectually huge groupings. Likewise it has been successfully used to perceive associations between individuals in friendly co-tasks, including relative status, precariousness, and the idea of cozy relationship.

3.7 *Classification Model:*

For these tasks, we train two Linear Support Vector Machines with TF-IDF weighted combinations of word and character n grams and LIWC features. Our character n gram features include all 2- to 4-grams; our word n gram features contain unigrams and bigrams; our LIWC features contain all the lexical indexes output by LIWC. We use a smoothed TF-IDF approach–implemented as $tf(t) \times log(\frac{N+1}{nt+1})$– where $tf(t)$ is the number of times the unigram or bigram t occurs, N is the number of documents and nt is the number of documents containing the unigram or bigram t. We limit our text prepossessing to sentence segmentation, tokenization, using a simple, social-media aware tokenizer, and ignoring case.

<u>MATLAB Code:</u>

```
setupFilePath = fileparts(mfilename('fullpath'));
speech2textFilePath = fileparts(setupFilePath);
addpath(speech2textFilePath);
savepath
% Install SpeechToText automation algorithm in Audio Labeler app. If Audio
% Labeler is already open, it needs to be restarted for the algorithm to
% show up.
if ~isempty(ver('audio')) && ~verLessThan('audio','2.1')
    registry = audio.labeler.automation.AutomationAlgorithmRegistry.getInst
    registry.addAlgorithm('SpeechToTextAutomation');
end
```

4. Results, Discussion and Analysis

Generally, data analysis and classification is recognized as a dual-step technique in which the initiative is the training stage, where a classification model develops the classifier by seeking from a dataset composed of databank tuples and are linked with labels. A label is nothing but a distinct value and each and every value serves as a class or category. The second and the most important step is to manipulate the model for classification and then the accuracy and efficiency of the classifier is predicted and estimated [9, 10, 11, and 12].

The average accuracy the sentiment model provides is nearly about 80%. There are multiple algorithms that are for mood prediction like SVM and Random Forest. Most of the papers mentioning the tweeter, Facebook and Instagram data has been used for sentiment analysis and the big tech company's used this big data and analyze this data to check the mood and emotions of the user towards a particular thing. To check the mood and opinions of people in 2020 from January to May on 'Metoo' hashtags and also from the women hashtags the Metoo there are 50429 tweets and for hashtags Women, there are about 50430 and most of the tweets wherein January and the lowest number in a tweet for both is on February[19].

As mentioned above the Python programming language best fits for data analysis and categorization so, for Sentimental Analysis, data can be classified and analyzed using different Machine learning algorithms and techniques. The method for given sentimental analysis is the Random Forest Classifier of Machine learning [9, 10, 11, and 12].

The research containing in this paper reflects the general issue of today's world, which is totally about the views of women. The data for research work is extracted from the insights of the tweets to analyzes, how does a woman feel comfortable discussing her views, approaches, difficulties, and two specific hashtags "#Women" & " #Meeto" using tweets that will be conducted in English.

The research work will be saved in a CSV file, containing the data withdrawing which is processed through the Twitter API and its online servers protected in the programing language of python. Machine learning uses mathematical and statically methods to refine the data. There are various platforms to work on but Jupyter notebook was found to be a more powerful and interactive tool to use. The useful built-in libraries and various demanding packages needed to be installed earlier for working on this project. The data of tweets were taken in a period of 2020 Jan till 2020 may for (#women) and (#meeto) respectively for the sentimental analyses of the public which will be stored in CSV extension [21].

Sentiment Analysis mainly focuses on the Behavior/mood of an abject the goal is to create a model that can understand the mood. The following screenshot Figure 5 defines the output of a test file that we give to him for testing but this is only possible when the code is already passed the testing and training process the dataset is about 10314 rows all are different from one another this is the dataset that helps us in training and testing.

```
       Index  ... label (depression result)
0        106  ...                          0
1        217  ...                          0
2        220  ...                          0
3        288  ...                          0
4        540  ...                          0
...      ...  ...                        ...
10309  802309  ...                         1
10310  802310  ...                         1
10311  802311  ...                         1
10312  802312  ...                         1
10313  802313  ...                         1

[10314 rows x 3 columns]
```

Figure 4: (Dataset of Sentiment Analysis)

We have taken the 20% of the dataset elements for testing and 80% of dataset elements for training. The code is using the testing data cells randomly which you can see in the given Figure 6 below. The output as anybody can see creates the confusion matrix that also known as the error matrix and its shows a summary of your problems or prediction results. The number is correct and incorrect predictions.

After importing the main libraries we given the dataset to the code dropna function is used to drop any null values in the dataset and after that we fetch the info of the updated dataset stop words is used to eliminated the common word that are present in the dataset. The Porter stemmer is function of NLTK that helps in producing morphological variants of a root then word embedding come into place after that we implement RNN algorithm and the model will be a sequential model after this step the training dataset will spilt into testing and training sets. In this testing set is 20% and the training sets are 80%. The model will be ready after training now we can make a confusion matrix and we will check the accuracy, F1 score, precision and Recall score.

```
Accuracy of RNN classifier = 0.9946679592825982

F1 Score of RNN classifier = 0.994670193585819

Precision of RNN classifier = 0.9946731240651314

Recall of RNN classifier = 0.9946679592825982
```

Figure 5: (Scores of RNN Algorithm)

Now implement random forest and fit the model after this step the plotting is started the heatmap is generated with the help of the dataset the heatmap graph will help in labelling the Actual and predicted values into the matrix.

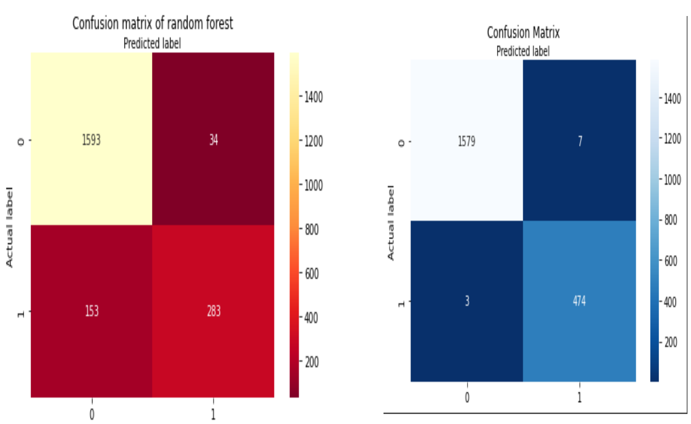

Figure 6: (Confusion matrix of Random forest and RNN

4.1 Application of Random Forest on the given model

The dataset consists of more than ten thousand comments or statements out of which more than half are optimistic, and the rest are aggressive. If a comment or statement contains more than 20% English that it will not be considered in the dataset as this study mainly focuses on Roman Urdu comments or statements (Duyu and Furu, 2014; Dimitris Effrosynidis). In this model, a Random forest classifier randomly allocates the data into training and testing halves. On execution, the forest classifiers are contoured with two arrays, one containing training data and the other contains the target values of the testing data (Duyu and Furu, 2014; Dimitris Effrosynidis).

4.2 Analysis of Data in Sentiment Analysis

Raw data is analyzed by applying NLP (Natural Language Processing), and it helps you in multiple domains like improving processes plus helps in decision-making, customer mindset for a specific product, and much more. In this code, we are checking if the object of a person is in a good or bad mood using a Random forest classifier. The data is given in a CVS file and contains different types of conversation/phrases to see their moods. The code is analyzing the conversation in the dataset by using Random Forest the algorithm uses randomness for building each tree. Afterward, the algorithm uses the tree prediction power to make a correct prediction. For this, the accuracy of the algorithm is 75%. The limitation of this algorithm when the tree is too large which makes it slow in real-time predictions.

The following steps are followed in the analysis of the dataset:

- Random Forest implements Bootstrapping to the datasets.
- It trains the dataset on each of the bootstrapped dataset elements independently.
- Random Forest don't use every feature for training purposes it just randomly select a subset of feature for each tree and use only them for training.
- After these sets the random forest builds the decision tree.
- All the trees are different from one another.

- Now we take a new data point and pass it through each tree.
- Now the trees are all set the dataset are starting to train every tree to generate the answer and the algorithm combine these answer.
- The random features selection helps to reduce the correlation of trees.

4.3 Discussion:

First of all, we import all the mandatory libraries like NumPy, pandas, seaborn, sklearn, etc. The dataset is set in the code to train the dataset is in sentences so it is hard for the program to learn from the whole sentence so for splitting the sentences into words we used the function named word_tokenize we tokenize Natural language with the help of (nltk) library the dataset will come in rough form so for the code to learn we used porter stemmer and onehot encoder. After the splitting of the words, the model starts to implement Random Forest Classifier and RNN on the split data and apply fitting for x_train and y_train, and eventually print it in a confusion matrix. The score and accuracy shows in the text form in the terminal. There are several other papers in which they are working on conversational data using sentiment analysis. One of the famous study subjects in current NLP is the texts corpora. Sentiment evaluation has its roots in herbal language processing and linguistics. Sentiment analysis tools and techniques by analyzing recorded voice communication data. Results show that the strategies that may be applied give smart classification results, though most sentences are neutral, mostly affirmative once expressing feelings. We have used a conversational dataset and tokenize it and then show the mood of an object based on the conversation [17].

5. Recommendations

The study evaluated the reason underpinning the ever-increasing suicidal and depression posts posted by users on social media platforms. It is explored from the findings that multiple social media users actively engage in sad posts and display their anger and anxiety constantly that impact their mental health conditions. Unfortunately, though, Pashto and Urdu are widely being spoken around the globe, with nearly 50-60 million dialects in Asia (Dinakhel, 2020). While, due to the use of multiple languages, it has become challenging to examine user sentiments and provide accurate insights for the mental health sector. Therefore, the use of machine learning models deployed and studied in the present study to evaluate stress and depression symptoms by combining lexicon of multilingual texts for assessing polarity of each word in sentiment analysis approach.

Though, the study finding showed accurate results by using lexicon-based sentiment analysis, which mainly focuses on extracting data based on the datasets and corpus built to depict the meaning of keywords of user's tweets and help define their polarity score. However, it is one of the constraints as it is studied that some of the incorrect scorings of the opinion of words are performed by existing lexicons, which was SentiWordNet (Baccianella et al., 2010). Therefore, it is recommended that future studies can focus on including domain-specific vocabulary to mitigate the existing issue and improve the efficacy of sentiment classification.

In this study, a random forest algorithm was also applied alongside the Lexicon approach due to the hybrid model applied. Although, the experiment performed to depict statistical results in determining polarity showed better outcomes and efficient results. It is also evaluated that the Random Forest algorithm possesses the strength of its unique ability to calculate several feature/variable importance measures, which could be the basis for increasing its explainability (Moreno-Marco et al., 2018). However, regardless of the fact that the random forest algorithm performed better, it lacked in certain features, comprising certain limitations. For instance, it was identified that a large number of trees make the algorithms too slow and inefficient for real-time predictions.

Therefore, the use of neural networks and deep learning models can be applied in detecting sentiments and equally quantifying the polarity of each word. To support this, He and Cao (2018) applied a deep convolutional neural network to learn deep-learned features and raw speech waveforms for depression analysis. This approach reflected the robustness and effectiveness in diagnosing depression and stress based on the databases comprising words. Thus, this approach can be applied and proposed as a sufficient machine learning technique to analyse users' words, texts, and opinions in stress and depression by using sentiment analysis.

The future work can also focus on speech-based information to analyse users' sentiments by crawling and detecting Facebook or Twitter API to realise the sentiments and opinions of individuals in terms of depression and stress. Besides, this can be a feasible technique to depict statistical technique for defining polarity of each word and match the relevance of score to other Python and NLP libraries for categorizing it into positive, negative and neutral.

6. Future Implications

In terms of future work, the current system may be be upgraded to other social media platforms such as Instagram and Facebook to check all social media activity to assess user mental health state. The extensive information from multiple social media platforms can help the health sector and specific health regulatory bodies to assess users' mental health conditions via their social media posts, comments, and shared opinion. Likewise, the implementation can also be used check online activity in close environment such as rehab clinic to monitor the progress of patients and users in real-time. Such a technique can easily scrape keywords based on users' online activity, which will help medical professionals assess the quality of care for patients in a better way.

Furthermore, the proposed mode has focused on the attributes of a hybrid model, where the use of a machine learning model can quickly help measure and statistically analyse users' mental state condition by assessing the magnitude of tweets. The overall model will provide the IT students and professionals to realize the implication of algorithms in medical-related conditions and how the local languages can be easily used for detecting stress and depression in a multilingual context. This will enable IT professionals and scientists to realize the benefit of sentiment analysis models in assessing users' health conditions and also closely monitoring the use of multiple state-of-the-art algorithms to predict its accuracy and efficacy in depicting results.

7. Conclusion

The part of the paper entailed a detailed overview of the generalized statements determining the overall results regarding the detection of stress and depression using the sentiment analysis approach. It further provided a better understanding of the implication of lexicon-based approach and random forest classifier as a statistical approach to develop a hybrid model for accurate and precise results in terms of categorizing words into positive, negative and neutral. Finally, the overall results and findings were summarized, and the limitations were also highlighted on which specific recommendations were proposed reflecting the use of Neural Network and API for extracting data and training models.

References

Ahmad, H., Asghar, M.Z., Alotaibi, F.M. and Hameed, I.A., 2020. Applying deep learning technique for depression classification in social media text. *Journal of Medical Imaging and Health Informatics*, *10*(10), pp.2446-2451.

Alahmary, R. and Al-Dossari, H., 2021. A semiautomatic annotation approach for sentiment analysis. *Journal of Information Science*, p.01655515211006594.

AlSagri, H.S. and Ykhlef, M., 2020. Machine learning-based approach for depression detection in twitter using content and activity features. IEICE Transactions on Information and Systems, 103(8), pp.1825-1832.

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H. and Shah, J., 2020. Sentiment analysis of extremism in social media from textual information. *Telematics and Informatics*, *48*, p.101345.

Baid, P., Gupta, A. and Chaplot, N., 2017. Sentiment analysis of movie reviews using machine learning techniques. *International Journal of Computer Applications*, *179*(7), pp.45-49.

Bakshi, R.K., Kaur, N., Kaur, R. and Kaur, G., 2016, March. Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)* (pp. 452-455). IEEE.

Budiyanto, S., Sihombing, H.C. and Rahayu, F.I., 2019. Depression and anxiety detection through the Closed-Loop method using DASS-21. *Telkomnika*, *17*(4), pp.2087-2097.

Cacheda, F., Fernandez, D., Novoa, F.J. and Carneiro, V., 2019. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research*, *21*(6), p.e12554.

Cacheda, F., Fernandez, D., Novoa, F.J. and Carneiro, V., 2019. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research*, *21*(6), p.e12554.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. and Zdeborová, L., 2019. Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), p.045002.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. and Zdeborová, L., 2019. Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), p.045002.

Cheerkoot-Jalim, S. and Khedo, K.K., 2020. A systematic review of text mining approaches applied to various application areas in the biomedical domain. *Journal of Knowledge Management*.

Mašanović, B., Popović, S., Bjelica, D., Vukotic, M. and Zoric, G., 2018. The effects of physical activity on depressive symptoms among elderly people: a systematic review. In *Proceedings World Congress of Performance Analysis of Sport XII* (Vol. 423).

Garg, R., Oh, E., Naidech, A., Kording, K. and Prabhakaran, S., 2019. Automating ischemic stroke subtype classification using machine learning and natural language processing. *Journal of Stroke and Cerebrovascular Diseases*, *28*(7), pp.2045-2051.

Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B. and Ziv, G., 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, *38*, p.100958.

Zech, J., Pain, M., Titano, J., Badgeley, M., Schefflein, J., Su, A., Costa, A., Bederson, J., Lehar, J. and Oermann, E.K., 2018. Natural language–based machine learning models for the annotation of clinical radiology reports. *Radiology*, *287*(2), pp.570-580.

Trappey, A.J., Trappey, C.V., Wu, J.L. and Wang, J.W., 2020. Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Advanced Engineering Informatics*, *43*, p.101027

Chen, Y., Zheng, W., Li, W. and Huang, Y., 2021. Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognition Letters*, *144*, pp.1-5.

Iwendi, C., Bashir, A.K., Peshkar, A., Sujatha, R., Chatterjee, J.M., Pasupuleti, S., Mishra, R., Pillai, S. and Jo, O., 2020. COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health*, *8*, p.357.

Ng, Q.X., Peters, C., Ho, C.Y.X., Lim, D.Y. and Yeo, W.S., 2018. A meta-analysis of the use of probiotics to alleviate depressive symptoms. *Journal of affective disorders*, *228*, pp.13-19.

Putnick, D.L., Sundaram, R., Bell, E.M., Ghassabian, A., Goldstein, R.B., Robinson, S.L., Vafai, Y., Gilman, S.E. and Yeung, E., 2020. Trajectories of maternal postpartum depressive symptoms. *Pediatrics*, *146*(5).

Edgcomb, J.B. and Zima, B., 2019. Machine learning, natural language processing, and the electronic health record: innovations in mental health services research. *Psychiatric Services*, *70*(4), pp.346-349.

Le Glaz, A., Haralambous, Y., Kim-Dufor, D.H., Lenca, P., Billot, R., Ryan, T.C., Marsh, J., Devylder, J., Walter, M., Berrouiguet, S. and Lemey, C., 2021. Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, *23*(5), p.e15708.

Singh-Manoux, A., Dugravot, A., Fournier, A., Abell, J., Ebmeier, K., Kivimäki, M. and Sabia, S., 2017. Trajectories of depressive symptoms before diagnosis of dementia: a 28-year follow-up study. *JAMA psychiatry*, *74*(7), pp.712-718.

Harshfield, E.L., Pennells, L., Schwartz, J.E., Willeit, P., Kaptoge, S., Bell, S., Shaffer, J.A., Bolton, T., Spackman, S., Wassertheil-Smoller, S. and Kee, F., 2020. Association between depressive symptoms and incident cardiovascular diseases. *JAMA*, *324*(23), pp.2396-2405.

Twenge, J.M. and Joiner, T.E., 2020. US Census Bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. *Depression and anxiety*, *37*(10), pp.954-956.

Köhler-Forsberg, O., N. Lydholm, C., Hjorthøj, C., Nordentoft, M., Mors, O. and Benros, M.E., 2019. Efficacy of anti-inflammatory treatment on major depressive disorder or depressive symptoms: meta-analysis of clinical trials. *Acta Psychiatrica Scandinavica*, *139*(5), pp.404-419.

Fullana, M.A., Hidalgo-Mazzei, D., Vieta, E. and Radua, J., 2020. Coping behaviors associated with decreased anxiety and depressive symptoms during the COVID-19 pandemic and lockdown. *Journal of Affective Disorders*, *275*, pp.80-81.

Huang, Y. and Zhao, N., 2020. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. *Psychiatry research*, *288*, p.112954.

Keyes, K.M., Gary, D., O'Malley, P.M., Hamilton, A. and Schulenberg, J., 2019. Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. *Social psychiatry and psychiatric epidemiology*, *54*(8), pp.987-996.

Cheng, Q., Li, T.M., Kwok, C.L., Zhu, T. and Yip, P.S., 2017. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. *Journal of medical internet research*, *19*(7), p.e243.

Chiba, S., Numakawa, T., Ninomiya, M., Richards, M.C., Wakabayashi, C. and Kunugi, H., 2012. Chronic restraint stress causes anxiety-and depression-like behaviors, downregulates glucocorticoid receptor expression, and attenuates glutamate release induced by brain-derived neurotrophic factor in the prefrontal cortex. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *39*(1), pp.112-119.

Chilwal, B. and Mishra, A.K., 2021. Extraction of Depression Symptoms From Social Networks. *The Smart Cyber Ecosystem for Sustainable Development*, p.307.

Cuijpers, P. and Karyotaki, E., 2021. The effects of psychological treatment of perinatal depression: an overview. Archives of Women's Mental Health, pp.1-6.

De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E., 2013, June. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

Deshpande, M. and Rao, V., 2017, December. Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 858-862). IEEE.

Deshpande, M. and Rao, V., 2017, December. Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 858-862). IEEE.

Deshpande, M. and Rao, V., 2017, December. Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)* (pp. 858-862). IEEE.

Dey, S.H.A.R.M.I.S.T.H.A., Sarkar, I., Chakraborty, S.R.A.B.A.N.T.I. and Roy, S.A.N.D.I.P., 2020. DEPRESSION DETECTION USING INTELLIGENT ALGORITHMS FROM SOCIAL MEDIA CONTEXT-STATE OF THE ART, TRENDS AND FUTURE ROADMAP. *Xi'an DianziKejiDaxueXuebao/Journal of Xidian University· August*.

Dey, S.H.A.R.M.I.S.T.H.A., Sarkar, I., Chakraborty, S.R.A.B.A.N.T.I. and Roy, S.A.N.D.I.P., 2020. DEPRESSION DETECTION USING INTELLIGENT ALGORITHMS FROM SOCIAL MEDIA CONTEXT-STATE OF THE ART, TRENDS AND FUTURE ROADMAP. *Xi'an DianziKejiDaxueXuebao/Journal of Xidian University· August*.

Dinakhel, M.A., 2020. Comparison of Two Dialects of Pashto, Spoken in Afghanistan and Pakistan: Kandahari Dialect and Yusufzai Dialect. *Central Asia*, *86*(Summer), pp.127-139.

Dreisbach, C., Koleck, T.A., Bourne, P.E. and Bakken, S., 2019. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International journal of medical informatics*, *125*, pp.37-46.

Fernández, E.M. and Cairns, H.S., 2011. *Fundamentals of psycholinguistics*. John Wiley & Sons.

Gaind, B., Syal, V. and Padgalwar, S., 2019. Emotion detection and analysis on social media. *arXiv preprint arXiv:1901.08458*.

Ghosh, C.C., McVicar, D., Davidson, G. and Shannon, C., 2021. Measuring diagnostic heterogeneity using text-mining of the lived experiences of patients. *BMC psychiatry*, *21*(1), pp.1-12.

Gonzalez-Hernandez, G., Sarker, A., O'Connor, K. and Savova, G., 2017. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearbook of medical informatics*, *26*(01), pp.214-227.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. Machine learning basics. Deep learning, 1(7), pp.98-164.

Gundecha, P. and Liu, H., 2012. Mining social media: a brief introduction. *New directions* *in informatics, optimization, logistics, and production*, pp.1-17.

Hammen, C.L., 2015. Stress and depression: old questions, new approaches. *Current Opinion in Psychology*, *4*, pp.80-85.

Hasan, A.N., Twala, B. and Marwala, T., 2014, July. Moving towards accurate monitoring and prediction of gold mine underground dam levels. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 2844-2849). IEEE.

Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H. and Ulhaq, A., 2018. Depression detection from social network data using machine learning techniques. Health information science and systems, 6(1), pp.1-12.

Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H. and Ulhaq, A., 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems*, *6*(1), pp.1-12.

Jalonen, H., 2014. Social media–an arena for venting negative emotions. *Online Journal of Communication and Media Technologies*, *4*(October 2014-Special Issue), pp.53-70.

Jamil, Z., 2017. *Monitoring tweets for depression to detect at-risk users* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp.255-260.

Kenyon-Dean, K., Ahmed, E., Fujimoto, S., Georges-Filteau, J., Glasz, C., Kaur, B., Lalande, A., Bhanderi, S., Belfer, R., Kanagasabai, N. and Sarrazingendron, R., 2018, June. Sentiment analysis: It's complicated!. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 1886-1895).

Khan, F.H., Qamar, U. and Bashir, S., 2017. A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems*, *51*(3), pp.851-872.

Khoo, C.S. and Johnkhan, S.B., 2018. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, *44*(4), pp.491-511.

Kotozaki, Y., 2020. An Exploratory Descriptive Study: Emotional Extraction of Postpartum Women Who Engaged in Horticulture Activities by Text Mining. *Psychology*, *11*(10), pp.1481-1492.

Kubick, N., Pajares, M., Enache, I., Manda, G. and Mickael, M.E., 2020. Repurposing zileuton as a depression drug using an AI and in vitro approach. *Molecules*, *25*(9), p.2155.

Kundi, F.M., Ahmad, S., Khan, A. and Asghar, M.Z., 2014. Detection and scoring of internet slang for sentiment analysis using SentiWordNet. *Life Science Journal*, *11*(9), pp.66-72.

Kurnia, P.F., 2018. Business intelligence model to analyze social media information. *Procedia Computer Science*, *135*, pp.5-14.

Lanzone, J., Cenci, C., Tombini, M., Ricci, L., Tufo, T., Piccioli, M., Marrelli, A., Mecarelli, O. and Assenza, G., 2020. Glimpsing the impact of COVID19 lock-down on people with epilepsy: a text mining approach. *Frontiers in Neurology*, *11*.

Lee, S.P., Sung, I.K., Kim, J.H., Lee, S.Y., Park, H.S. and Shim, C.S., 2015. The effect of emotional stress and depression on the prevalence of digestive diseases. *Journal of neurogastroenterology and motility*, *21*(2), p.273.

Li, G., Li, B., Huang, L. and Hou, S., 2020. Automatic Construction of a Depression-Domain Lexicon Based on Microblogs: Text Mining Study. *JMIR medical informatics*, *8*(6), p.e17650.

Li, R. and Liu, Z., 2020. Stress detection using deep neural networks. *BMC Medical Informatics and Decision Making*, *20*(11), pp.1-10. William, D. and Suhartono, D., 2021. Text-based Depression Detection on Social Media Posts: A Systematic Literature Review. *Procedia Computer Science*, *179*, pp.582-589.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machine learning in agriculture: A review. Sensors, 18(8), p.2674.

Lin, Z., 2021. A Methodological Review of Machine Learning in Applied Linguistics. *English Language Teaching*, *14*(1), pp.74-85.

Losada, D.E. and Crestani, F., 2016, September. A test collection for research on depression and language use. In *International Conference of the Cross-Language Evaluation Forum for European Languages* (pp. 28-39). Springer, Cham. Trotzek, M., Koitka, S. and Friedrich, C.M., 2017, September. Linguistic Metadata Augmented Classifiers at the CLEF 2017 Task for Early Detection of Depression. In *CLEF (Working Notes)*.

Ma, L., Wang, Z. and Zhang, Y., 2017, May. Extracting depression symptoms from social networks and web blogs via text mining. In *International Symposium on Bioinformatics Research and Applications* (pp. 325-330). Springer, Cham.

Marx, P., Antal, P., Bolgar, B., Bagdy, G., Deakin, B. and Juhasz, G., 2017. Comorbidities in the diseasome are more apparent than real: what Bayesian filtering reveals about the comorbidities of depression. *PLoS computational biology*, *13*(6), p.e1005487.

Meitei, K.T. and Singh, H.S., 2019. Prevalence of Depression Among Indian Population. *Online J Health Allied Scs*, *18*(4), p.12.

Monner, D., Vatz, K., Morini, G., Hwang, S.O. and DeKeyser, R., 2013. A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition*, *16*(2), pp.246-265.

Moreno, M.A., Jelenchick, L.A., Egan, K.G., Cox, E., Young, H., Gannon, K.E. and Becker, T., 2011. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, *28*(6), pp.447-455.

Nadeem, M., 2016. Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384*.

Nadeem, M., 2016. Identifying depression on Twitter. *arXiv preprint arXiv:1607.07384*.

Orabi, A.H., Buddhitha, P., Orabi, M.H. and Inkpen, D., 2018, June. Deep learning for depression detection of twitter users. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic (pp. 88-97).

Pahwa, B., Taruna, S. and Kasliwal, N., 2018. Sentiment analysis-strategy for text preprocessing. *Int. J. Comput. Appl*, *180*, pp.15-18.

Park, A., Conway, M. and Chen, A.T., 2018. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior*, *78*, pp.98-112.

Patel, F., Thakore, R., Nandwani, I. and Bharti, S.K., 2019, December. Combating depression in students using an intelligent chatBot: a cognitive behavioral therapy. In 2019 IEEE 16th India Council International Conference (INDICON) (pp. 1-4). IEEE.

Paul, S., Jandhyala, S.K. and Basu, T., 2018, August. Early Detection of Signs of Anorexia and Depression Over Social Media using Effective Machine Learning Frameworks. In *CLEF (Working notes)*.

Peng, Z., Hu, Q. and Dang, J., 2019. Multi-kernel SVM based depression recognition using social media data. *International Journal of*

Machine Learning and Cybernetics, *10*(1), pp.43-57.

Picardi, A., Lega, I., Tarsitani, L., Caredda, M., Matteucci, G., Zerella, M.P., Miglio, R., Gigantesco, A., Cerbo, M., Gaddini, A. and Spandonaro, F., 2016. A randomised controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care. *Journal of affective disorders*, *198*, pp.96-101.

Plonsky, L. and Ghanbar, H., 2018. Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, *102*(4), pp.713-731.

Pradha, S., Halgamuge, M.N. and Vinh, N.T.Q., 2019, October. Effective text data preprocessing technique for sentiment analysis in social media data. In *2019 11th international conference on knowledge and systems engineering (KSE)* (pp. 1-8). IEEE.

Rambocas, M. and Gama, J., 2013. *Marketing research: The role of sentiment analysis* (No. 489). Universidade do Porto, Faculdade de Economia do Porto.

Regier, D.A., Kuhl, E.A. and Kupfer, D.J., 2013. The DSM‐5: Classification and criteria changes. *World psychiatry*, *12*(2), pp.92-98.

Richter, T., Fishbain, B., Markus, A., Richter-Levin, G. and Okon-Singer, H., 2020. Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Scientific reports*, *10*(1), pp.1-12.

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š. and Sedlmair, M., 2018. More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, *12*(2-3), pp.140-157.

Sadeque, F., Xu, D. and Bethard, S., 2018, February. Measuring the latency of depression detection in social media. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 495-503).

Salina, A., 2015, September. Business reviews classification using sentiment analysis. In *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (pp. 247-250). IEEE.

Sau, A. and Bhakta, I., 2019. Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*, *16*, p.100228.

Schetter, C.D. and Tanner, L., 2012. Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice. *Current opinion in psychiatry*, *25*(2), p.141.

Settles, B., T LaFlair, G. and Hagiwara, M., 2020. Machine learning–driven language assessment. *Transactions of the Association for computational Linguistics*, *8*, pp.247-263.

Seyeditabari, A., Tabari, N. and Zadrozny, W., 2018. Emotion detection in text: a review. *arXiv preprint arXiv:1806.00674*.

Sharma, N., Pabreja, R., Yaqub, U., Atluri, V., Chun, S.A. and Vaidya, J., 2018, May. Web-based application for sentiment analysis of live tweets. In Proceedings of the 19th Annual International Conference on Digital government research: Governance in the data Age (pp. 1-2).

Sharmin, S. and Zaman, Z., 2017, December. Spam detection in social media employing machine learning tool for text mining. In *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 137-142). IEEE.

Shen, C.W., Chen, M. and Wang, C.C., 2019. Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computers in Human Behavior*, *101*, pp.474-483.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.S. and Zhu, W., 2017, August. Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. In IJCAI (pp. 3838-3844).

Shi, L.L., Liu, L., Wu, Y., Jiang, L. and Hardy, J., 2017. Event detection and user interest discovering in social media data streams. *IEEE access*, *5*, pp.20953-20964.

Smirnova, D., Cumming, P., Sloeva, E., Kuvshinova, N., Romanov, D. and Nosachev, G., 2018. Language patterns discriminate mild depression from normal sadness and euthymic state. *Frontiers in Psychiatry*, *9*, p.105.

Stanton, R., To, Q.G., Khalesi, S., Williams, S.L., Alley, S.J., Thwaite, T.L., Fenning, A.S. and Vandelanotte, C., 2020. Depression, anxiety and stress during COVID-19: associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. *International journal of environmental research and public health*, *17*(11), p.4065.

Stappen, L., Cummins, N., Meßner, E.M., Baumeister, H., Dineley, J. and Schuller, B., 2019, May. Context modelling using hierarchical attention networks for sentiment and self-assessed emotion detection in spoken narratives. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6680-6684). IEEE.

Suciati, S., 2021, June. Favourite Artificial Intelligence-based speaking applications for autonomous learning. In *UNNES-TEFLIN National Seminar* (Vol. 4, No. 1, pp. 205-214).

Symeonidis, S., Effrosynidis, D. and Arampatzis, A., 2018. A comparative evaluation of preprocessing techniques and their interactions for twitter sentiment analysis. *Expert Systems with Applications*, *110*, pp.298-310.

Tran, B.X., Ha, G.H., Nguyen, D.N., Nguyen, T.P., Do, H.T., Latkin, C.A., Ho, C.S. and Ho, R.C., 2020. Global mapping of interventions to improve quality of life of patients with depression during 1990–2018. *Quality of Life Research*, pp.1-11.

Tran, B.X., Ha, G.H., Nguyen, D.N., Nguyen, T.P., Do, H.T., Latkin, C.A., Ho, C.S. and Ho, R.C., 2020. Global mapping of interventions to improve quality of life of patients with depression during 1990–2018. *Quality of Life Research*, pp.1-11.

Trifan, A., Antunes, R., Matos, S. and Oliveira, J.L., 2020. Understanding depression from psycholinguistic patterns in social media texts. *Advances in Information Retrieval*, *12036*, p.402.

Trupthi, M., Pabboju, S. and Narasimha, G., 2017, January. Sentiment analysis on twitter using streaming API. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 915-919). IEEE.

Vajjala, S., 2018. Machine Learning and Applied Linguistics. *arXiv preprint arXiv:1803.09103*.

Van der Zanden, R., Curie, K., Van Londen, M., Kramer, J., Steen, G. and Cuijpers, P., 2014. Web-based depression treatment: Associations of clients′ word use with adherence and outcome. *Journal of affective disorders*, *160*, pp.10-13.

Vasyliuk, V., Shyika, Y. and Shestakevych, T., 2020. Information System of Psycholinguistic Text Analysis. In *COLINS* (pp. 178-188).

Wang, S.H., Ding, Y., Zhao, W., Huang, Y.H., Perkins, R., Zou, W. and Chen, J.J., 2016. Text mining for identifying topics in the literatures about adolescent substance use and depression. *BMC public health*, *16*(1), pp.1-8.

Wang, Y. and Berwick, R.C., 2013. Formal relational rules of english syntax for cognitive linguistics, machine learning, and cognitive computing. *Journal of Advanced Mathematics and Applications*, *2*(2), pp.182-195.

Wang, Y., Wang, Z., Li, C., Zhang, Y. and Wang, H., 2020. A Multitask Deep Learning Approach for User Depression Detection on Sina Weibo. *arXiv preprint arXiv:2008.11708*.

Wu, C.S., Kuo, C.J., Su, C.H., Wang, S.H. and Dai, H.J., 2020. Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. *Journal of affective disorders*, *260*, pp.617-623.

Xu, G., Meng, Y., Qiu, X., Yu, Z. and Wu, X., 2019. Sentiment analysis of comment texts based on BiLSTM. Ieee Access, 7, pp.51522-51532.

Xu, H., Phan, T.Q. and Tan, B., 2013. How does online social network change my mood? An empirical study of depression contagion on social network sites using text-mining.

Yadav, A. and Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), pp.4335-4385.

Yang, L., Li, Y., Wang, J. and Sherratt, R.S., 2020. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE Access, 8, pp.23522-23530.

Yang, X., McEwen, R., Ong, L.R. and Zihayat, M., 2020. A big data analytics framework for detecting user-level depression from social networks. International Journal of Information Management, 54, p.102141.

Zucco, C., Calabrese, B. and Cannataro, M., 2017, November. Sentiment analysis and affective computing for depression monitoring. In *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 1988-1995). IEEE.

Tripathy, A., Agrawal, A., and Santanu, R., 2015, "Classification of Sentimental Reviews Using Machine Learning Techniques", Proceedings of 3rd International Conference on Recent Trends in Computing, pp. 821-829.

Govindarajan, M., 2014, "Sentiment Analysis of Restaurant Reviews Using Hybrid Classification Method", International Journal of Soft Computing and Artificial Intelligence, Volume 2, pp. 330-344.

Luhn, H., 1958 "The Automatic Creation of Literature Abstracts", IBM Journal of Research and Development, pp. 159-165.

Field and John (2009), "Dividing listening into its components", Listening in the Language Classroom, Cambridge: Cambridge University Press, pp. 96–109, doi:10.1017/cbo9780511575945.008, I SBN 978-0-511-57594-5.

N. Bahrawi, 2019, "Online Realtime Sentiment Analysis Tweets by Utilizing Streaming API Features From Twitter," J. Penelit. Pos dan Inform., vol. 9, no. 1, pp. 53–62.

Wan, Y. and Gao, Q., 2015, November. An ensemble sentiment classification system of twitter data for airline services analysis. In *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 1318-1325). IEEE.

L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets using Naïve Bayes' and K-NN Classifier."

Niklas Donges, 2021, "A Complete Guide to the Random Forest Algorithm".

U. A. Siddiqua, T. Ahsan, and A. N. Chy, 2016, "Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog," in 2016 19th International Conference on Computer and Information Technology (ICCIT), pp. 304– 309.

G. Gautam and D. Yadav, 2014, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in 2014 Seventh International Conference on Contemporary Computing (IC3), pp. 437–442.

Duyu Tang and Furu Wei, 2014, developed a Deep Learning System, called "Coooll" which is a deep learning model that builds sentiment classifier from Tweets and manually annotated sentiment polarity.

Dimitris Effrosynidis (21st Conference on Theory and Practice of Digital Libraries).

Bilal, M., Israr, H., Shahid, M., and Khan, A., 2016, "Sentiment Classification of Roman-Urdu Opinions Using Naive Bayesian, Decision Tree and KNN Classification Techniques", Journal of King Saud University-Computer and Information Sciences, Volume 28, pp. 330-344.

Mukhtar, N., and Khan, M., 2017, "Urdu Sentiment Analysis Using Supervised Machine Learning Approach", International Journal of Pattern Recognition and Artificial Intelligence, Volume 32, pp. 1851001-1851015.

Hearst, M., 1992. Automatic acquisition of hyponyms from large text corpora. Proceedings of the 14th Conference on Computational Linguistics, (CCL' 92), Stroudsburg, PA, USA, pp: 539-545.

Khan, A.; Baharudin, B.; Khan, K. 2010, Efficient feature selection and domain relevance term weighting method for document classification. In Proceedings of the 2010 Second International Conference on the Computer Engineering and Applications (ICCEA), Bali Island, Indonesia, Volume 2, pp. 398–403

Litman, D. and Forbes, K., 2003, November. Recognizing emotions from student speech in tutoring dialogues. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721)* (pp. 25-30). IEEE.

Liu, B. 2010, "Sentiment Analysis and Subjectivity". Appeared in Handbook of Natural LanguageProcessing, Indurkhya, N. &Damerau, F.J. [Eds.].

Kaur, 2021 "Social Issues Sentiment Analysis using Python", "www.researchgate.net/publication/312176414_Sentiment_Analysis_in_Python_using_NLTK"

Parmar, H., Bhanderi, S. and Shah, G., 2014, July. Sentiment mining of movie reviews using Random Forest with Tuned Hyperparameters. In *International Conference on Information Science* (pp. 1-6).

Kaur and Sharma, October 2020. Social Issues Sentiment Analysis using Python: https://www.researchgate.net/publication/346962692_Social_Issues_Sentiment_Analysis_using_Python.