Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SENTIMENT-ANALYSIS TO DETECT EARLY DEPRESSIVE SYMPTOM IN BANGLA LANGUAGE FROM SOCIAL MEDIA: A REVIEW STUDY

[1]MD. HASIBUL HASSAN, [2]AZRINA KAMARUDDIN, [2]MASRAH AZRIFAH AZMI MURAD

Software Engineering and Information System, Faculty of Computer Science & Information Technology, University Putra Malaysia, Malaysia

E-mail: [1]gs56817@student.upm.edu.my, [2]{azrina, masrah}@upm.edu.my

ABSTRACT

Social media platforms hold a vast volume of raw data that has been posted by people in the forms of texts, images, audio and video. People use this medium to express their thoughts and opinions. As a result, the data can be captured, categorized, and analyzed using Sentiment Analysis approaches to identify users' behavior, customer's feedback or gauge public opinions. WHO reported that the numbers on existing mental health disorders are a troubling phenomenon. The identification of mental health can be detected using several data domains such as: sensors, text, structured data, and multi-modal system use. Several researches focus on specific public sentiments for example Malays, English, Arabic, Chinese and Korean. However, very little research was conducted about sentiment analysis approaches implementation in Bangla language. The purpose of this paper is to explain the knowledge gap and the proposed model by using Bangla language sentiment analysis. In this paper we have reviewed 50 articles from which 18 are listed here that have the most similarity with our research. The review shows that the mostly used method in Sentiment-analysis is Machine Learning in the field of Opinion-mining. Furthermore, we have identified another variable that can be included to improve the existing algorithm.

Keywords: *Early Depressive Symptom, Sentiment Analysis, Bangla Language, Social Media*

1. INTRODUCTION

In the era of modern technology, we live in an age of borderless boundaries where we received, created, shared, and stored within the borderless world where we are bombarded with lots of information. Social life has been brought too close by the dint of technical revolution. The vast use of social media has driven us successfully cross over distant places and share information with each other. Among them Facebook, Twitter, Instagram, Reddit etc. are the very common and highlighted names in the media-oriented world [1]. The users of social media are increasing by the day, especially Facebook. By this time it there are 2.77 billion users of Facebook worldwide [2]. In Facebook, people not only take the social opinion, they also express their personal feelings about some specific topics or matters. Facebook has become a vast medium of communication and sharing emotional data. Since people use Facebook as a way to compose their thoughts and share their opinions, we believe this may be an approach to recognize early symptoms of mental illness or depression [3], [4].

Depression is the most common, prevailing and growing mental illness. The long term effects of depression increasing [5]–[7]. The depressed persons are very much active in social media like Facebook as they are introvert in public [8]. It's very easy to analyze the personality and monitor the mental health issues. They post so much opinion and emotive data. This is helpful to express their feelings. More than that, it makes their posts easier to analyze.

Many studies on Sentiment-analysis are conducted on English Language [6], [9], [10]. Currently, there are researches on Sentiment-analysis that have crossed over the boundary of language and reaches all over the world. Some other languages have adopted Sentiment-analysis for various purposes like product review, advertisement, stock market analysis for investment etc. The other languages that have adopted Sentiment-analysis are Korean [11], Thailand [12], Arabic [13], Malay [14], Portuguese and Chinese [15].

Like other languages, the uses of Bangla language in Facebook are also increasing. In Bangladesh, people use Facebook on daily basis for their communication and business purposes. The use of

Facebook is not limited to these. Tt has evolved as the sharing medium where people share on a daily basis every single feeling or activity of theirs [16]. These posts contain both positive and negative emotive texts. These texts are direct representations of a person's hidden emotion especially the negative ones [5].
Sentiment-analysis is also known as emotion AI or opinion mining. Sentiment evaluation is a strategy that makes use of Natural Language Processing (NLP), computational linguistic and text evaluation to extract thoughts. Converting as well as deciphering opinion from texts to categorize them into normal, negative or positive emotions [17] are one of the remarkable contribution of Sentiment-analysis. Research on Sentiment-analysis has increased after a decade in 2004 [18]. Sentiment-analysis has been a preferred research approach not only in emotion detection but also in other sectors such as to update the quality of Airport Service [19]; to investigate the therapy of mental illness [20]; and to observe the business improvement in stock market [21]. Researchers divided Sentiment-analysis into three levels: sentence level; document level; and feature/aspect level [22]. Sentiment-analysis is performed in two ways namely Machine Learning Based and Lexicon Based. Machine Learning based approaches is mainly algorithm or classifier oriented whereas Lexicon based analysis is totally dependent on the linguistic dictionary [2].
Even in Bangla language several works have been done on opinion review and depression detection [9], [10], [16], [23]. Initially Sentiment-analysis was used for the opinion review for the E-commerce business and recording the public reaction on a particular product. In our study we used the Lexicon Based Analysis. The main aspect behind picking this approach is, in Bangladesh very few work have been done using Linguistic Dictionary [24]. One more thing, that in Lexicon Based Analysis, the overhead of training data is not considered as a major effect on accuracy [25].

The aims of this paper are: to understand the state-of-the-art methods of depression detection in sentiment analysis and identifying the research gap.

2. BACKGROUND STUDY

2.1. Social Media

The introduction of web 2.0 has changed the interaction with the social media world. Social media are used to communicate with the world. We share information and opinions with others. Social media are pivotal even in business communication, understanding, and improvement of their products as well as services by connecting worldwide. Social media users increased in numbers by the day. It is estimated that there will be 2.77 billion users worldwide by 2019 [2]. Social media has become a rich source of data that has been shared and uploaded as text, videos, photos, and audio [26]. Based on the uses of social media we have selected the major social media like Facebook, Twitter and Reddit.

2.2. Facebook

Social media is a goldmine of raw data that is unprocessed and used in the improvements in technology. Machine learning as well as artificial intelligence allows data to be analyzed and transformed into usable information that can support any business enterprises. [27]. Facebook, Twitter, etc. has become a great source of social communication medium now. Facebook is deemed a better place to express one's emotions briefly rather than twitter because of the post's character limit on Facebook (63206) is immensely larger than Twitter (280). Currently, Facebook is dealing with approximately 2 billion of users whereas Twitter has 319 million users. On the basis of using characters to express personal emotions briefly Facebook is the better choice [28]. Facebook becomes very popular amongst teenagers and young adults.
In Malaysia alone, the number of Facebook users has increased from 20.65 million to 24.85 million within 4 years (2017-2020) [29]. It is also forecasted that the number of users will increase up to 3 million to the total of user 27.34 million in 2025 [29]. Similarly in Singapore, the uprising use of Facebook is from 4.57 to 5.09 million users between the years of 2017-2025 [30]. In Thailand, the rate is 43.69 to 61.36 million between 2017 to 2025 [31]. In March 2020, there were 37, 912, 000 Facebook customers in Bangladesh. This accounted for 22.1% of its complete population. Majority of the user were men (71.6%). People aged 18 to 24 were the biggest user group (16,100,000). The best possible distinction between men and women was seen in the 25 to 34 age group where men lead by 6,800,000 [32].
As the achiever of mostly used social media in Bangladesh people are very much active on Facebook [33]. Although the initial motive of Facebook was communication but now it has become not only a business sector but also an opinion hub [34].

2.3. Twitter

Twitter is another social media popular for public opinion and comments. Hash tag comments are the trend of Twitter. It has become a popular medium for sharing information and personal feelings. However, character limitation to only 280 characters make it difficult to express any emotion precisely [2]. Twitter is the preferable social media for big data research rather than Facebook because of the availability of data [27]. Collecting data from Twitter is easier with the help of Twitter API [35].

2.4. Depression

Depression (major depressive disorder) is a common and serious medical illness that negatively influences how we experience the way we suppose and how we act. Depression deals feelings of disappointment and/or a loss of hobby in things to do that one once enjoyed [5]. Depression is a common occurrence nowadays in people of all ages, especially in the United States, where 6.7 percent of adults suffer from it, adversely impacting their thoughts and behaviors. Unipolar disorder affects nearly 24 percent of young adults (ages 20-24) and 18 percent of teens (ages 15-20) in Japan alone [36]. The worst case is depression leads to suicide. Every year almost 800000 individuals have committed suicide due to depression [4]. According to a World Health Organization survey from 2015, 300 million people are affected by depression, and the bad news is that the prevalence of depression among teenagers has risen in recent years from 8.7% to 11.3 percent, and among young adults from 8.8% to 9.6% [6]. Girls and women are at a greater risk than men for asthma, dementia, and silent stroke, with a 50 percent higher chance of being affected [6]. The above statistics are the motivational aspect of conducting this research. Though depression is a prevalent mental illness, the number of people that are taking treatment is inferior. They hide their emotions in their mind. That's why no symptoms come to light. Social media like Facebook, Twitter, etc. become very popular as they are free and safe, where they can use to share their daily life feelings on it and sought for others' help. That's why these social media posts' contents are valuable source of emotional information to analyze the presence of depression in them. The effect of depression and its increasing rate is discussed in section 2.5 briefly. Social media brings the whole world to a stone throw away in terms of communication. On the other hand people socialize so much on social media that they lost their own social morality [37] and they are becoming separated from each other [20]. This is one of the biggest reason behind sharing too much of negative or depressive posts on Facebook [38].

2.5. Social Media and Depression

When analyzing social media's relationship with mental health outcomes such as depression, findings reported mixed results. The relationship is complicated probably because it involves more than one psychological, social, behavioral and man or woman factors [39]. Whether social media is beneficial or damaging to mental health and well-being partly rely on the quality of social elements in the social community site's (SNS) surroundings [40]. While analyzing child and adolescent populations, research shows a steady relationship between cyber bullying and depression [41]. In 18-to-29-years-olds, it was found that advantageous social comparison is drastically negatively related with depressive symptoms [42]. A study showed expanded odds of depression in participants spending greater time on social media. Based on gratification theory, media used in a goal-directed way for individual gratification and satisfaction has similarities with addiction [43]. Griffiths defines addictive behavior as being characterized via six core elements of addiction: salience, temper modification, tolerance, withdrawal symptoms, conflict, and relapse and any behavior that fulfills these six criteria can be viewed as an addiction, consisting of social networking [44]. Further research is needed because of the high incidence of social media utilization and excessive rates of depression in young adults, with melancholy as the most common health problem for university students [45]. As in section 2.2, we saw growing rate of social media uses especially Facebook in Bangladesh. This correlate with a gradual increased in the effects of depression among the users by 52.2% [46].

2.6. Sentiment-Analysis

Sentiment-analysis is also known as emotion AI or opinion mining. Sentiment evaluation is a strategy that makes use of Natural Language Processing (NLP), computational linguistic and text evaluation to extract thoughts. Converting as well as deciphering opinion from texts to classify them into positive, negative, or natural sentiment [17] are among the remarkable contributions of Sentiment-analysis. Research on Sentiment-analysis has grown after a decade since

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2004 [18]. Sentiment-analysis is the preferred approach not only in emotion detection but also in other sectors for example to update the quality of Airport Service [19]; to investigate the therapy of mental illness [20]; to observe the business improvement in stock market [21], etc.

Researchers divided Sentiment-analysis into three levels: sentence-level; document-level; and feature/aspect-level [22].

2.6.1. Sentence-level

In the area of Sentiment-analysis, Sentence-level Sentiment-analysis takes major directions. The polarity of the sentences was measured and classified as having either positive, negative or neutral emotions hidden in the sentences based on the semantic information of the sentence.

For depression analysis Sentence-level analysis is being used, as it is directly collected from the users [2]. People share various post in social media which contains only a few sentences. Even though Twitter and Facebook have character restrictions, people still use Sentence to express emotions. That’s why Sentence-level is fit for analyzing the emotions of the text. This makes Sentence-level analysis very effective for analyzing the depressive symptoms.

2.6.2. Document-level

Document level analysis is performed on the given document. System reads the input data from the given document and categorized the sentences of that document into positive, negative and neutral statements. Here both sentence and document are accepted.

2.6.3. Feature/aspect-level

In this level of Sentiment-analysis aspects are extracted from the given text. Various algorithms are used to extract the sentiment, and for every aspect that sentiment is evaluated. Aspect level of Sentiment-analysis is also called entity/feature level analysis. Aspect level Sentiment-analysis is used to detect the presence of multiple sentimental entities in a single sentence.

2.7. Approaches of Sentiment-Analysis

Machine learning and Lexicon-based analysis are the two main approaches to Sentiment-analysis. Extraction and detection algorithms are performed to detect sentiment from data by Machine Learning approach while counting the positive and negative words from the relevant data is the primary working method of Lexicon-based analysis. Though many useful and accurate model in Sentiment-analysis were developed especially those focused on the English language, a current study brings it to light that some other models have been designed in other languages such as Korean [11], Thailand [12], Arabic [13], Malay [14], Portuguese and Chinese [15].

2.7.1. Lexicon-based approach

Lexicon-based analysis is also called unsupervised learning. This method doesn’t require any training data, and it only depends on the Lexicon-based dictionary. Most of the study adapted two methods while conducting research such as Sentiwordnet and TF-IDF. The occurrences of positive and negative words in the text data are calculated by this approach, like Sentiwordnet [17]. In TF-IDF, the words are converted into a number and calculated their frequency in terms of frequency-inverse documentation [2].

The effectiveness of Lexicon-based analysis depends on the resource and its quality. Lexicon based method doesn't need any data for training purpose. It's very fast with a small amount of data set. For better accuracy, we used a Lexicon-based database, including positive and negative words. The other advantage of this method is that it provides the counting of positive and negative words and flexibility with different languages. Avoiding the training data, the analysis has become very fast and significantly shows a better performance in terms of accuracy. Though lexicon-based analysis is an advanced method to analyze the emotional aspect of text posted by the users, still it's some drawback as well. This technique was designed for simple words used in the language. The complexity comes when it appears with slang, negation, and sarcasm [47]. The idea of using words to detect sentiment is not enough to conclude any decision strongly. Some problems that we faced with this approach were when some words contain different meanings in different applications; some sentences can express sentiment without using any sentiment words; or some may not express any sentiment while having sentiment words [48].

2.7.2. Machine-learning approach

Machine learning is known as supervised learning. This approach requires data for training purposes. SVM, RNN and Naive Baye’s are the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

most used model in this approach. For well-formed text, corpus Naïve Baye's is a successfully used method [49]. On the contrary, for low space, dataset SVM provides a good performance. To classify the random and non clustered data like Facebook, the performance of machine learning is very poor as the amount of data is huge, and it requires a huge training dataset [23], [48]. However, machine learning requires more time to perform because of the training data [14]. If the dataset is low, then the analysis will be faster, but it affects the accuracy to be a poorer classification [50]. The accuracy of machine learning depends on the performance of the classifiers. Some machine learning classifiers that are used in Sentiment-analysis are explained bellow:

2.7.2.1. Support-vector machines (SVM)

Support-vector machines in machine learning are known as supervised learning models. It is associated with learning algorithms that analyze data. Those data were used for classification and regression analysis. The Support Vector Machine (SVM) algorithm is a popular computing device studying device that gives solutions for each classification and regression problems. Though SVM has a milestone in the field of Machine Learning, it still has some limitations as the following;

- SVM can't handle large data sets.
- SVM is not suitable for noisy and overlapping data.
- The performance of SVM becomes poor if the number of features for each data point overflows the number of training data sample.

2.7.2.2. Naive baye's

Naïve Baye's is one of Machine Learning supervised algorithm. The Baye's theorem of probability is used by Navie Baye's algorithm. In Navie Baye's classifier the features are assumed statistically independent. The reliability of the classifier depends on the assumption of the independency of variables. Regardless of this assumption, it has proven itself to be a classifier with good results. Though it's a fast and simple classifier it's some drawbacks too.

- The accuracy of this classifier depends on the assumption of the independent variables. But in real life it's impossible to find any variables that are totally independent.
- If any mismatch happens between the training and test dataset during categorizing, then Naïve Baye's algorithm puts a zero probability. This is called a zero frequency. Zero-frequency hinder the model's ability to make any prediction.

2.7.2.3. Recurrent neural network (RNN)

Recurrent neural network (RNN) is an artificial neural network. In RNN a temporal sequence of nodes of a directed graph are connected along with each other. This allows RNN to store data for a short time. Derived from feed ahead neural networks, RNNs can use their internal kingdom (memory) to system variable size sequences of inputs. In both [9], [16] Machine Learning with Recurrent Neural Network (RNN) has been introduced in the field of Sentiment-analysis. They were focused on Bengali online sports newspaper Prothom-alo to analyze the review comments of the sports fan in Bangladesh. They collected the comments and tagged them as sentimental comments. Deep learning had been implemented, which is a subfield of machine learning [9]. And RNN was used as a classifier to categorize the data. In RNN training data are readily stored. This algorithm had been trained up with the input data set until it is able to predict the upcoming event based on the previous event during the training session. As RNN is developed to remember the events with the duration of time, RNN predicts events easily. The uniqueness of [9] is, it has been done on both Bengali and Hindi dataset. In [16], authors' focused on depression detection. Though RNN has the ability to store a bunch of data for prediction, it is not able to keep track of long term dependencies. That's why it's called Long Short Term Memory (LSTM). As the input dataset is a very long sequence of emotional texts, RNN is not suitable for this. RNN has better accuracy for short term dataset.

For a small amount of dataset, Machine learning performs good accuracies [16]. But the accuracy will drop when the dataset is bigger. This statement has also been supported by the authors [9]. All the works on social media done previously including the analysis on Bengali Language, were based on sentence level Sentiment-analysis only. The given input data was social media posts which are purely texts.

Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. REVIEWED ARTICLES

18 out of 50 papers were picked from the selection process after reading the original and in-depth scanning of the report. The papers were selected based on the most similarity with the purpose of the study, which stands for identifying gap of existing work in Bangla Language. These papers are collected from the reliable database like Emerald Insight, ACM, Science Direct, IEEE Xplore and Scopus. The reviewed Articles are divided into sub category as Lexicon-Based Approach and Machine Learning Based Approach. Table 1 is the representation of Lexicon-Based Approach category.

Table 1: Reviewed Articles on Lexicon-Based Approach

Author & Year	Title	Scope	Parameter	Language	Contribution
R. C. Dey et al, 2019 [24]	Sentiment analysis on Bengali text using lexicon based approach	Depression Detection	Text	Bangla	Linguistic Dictionary in Bangla Language
C. Dhaoui et al, 2017 [50]	Social media sentiment analysis: Lexicon versus Machine Learning	Opinion Review	Text	English	Comparison between Lexicon-based and Machine Learning Approach
X. Tao et al, 2016 [7]	Sentiment analysis for depression detection on social networks	Depression Detection	Text, Corpus	English	Use a depressive sentiment knowledge base and an algorithm to analyze textual data for depression detection.
D. Doğru et al, 2018 [51]	Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach	Opinion Review	Text	Turkish	To classify the sentiments according to the sentiment density in Turkish Language
S. Yuliyanti et al, 2017 [52]	Sentiment mining of community development program evaluation based on social media	Opinion Review	Text	English	Opinion review based on comments
M. Itani et al, 2017 [13]	Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media	Opinion Review	Text	Arabic	Sentiment-analysis to find out the mental state in Arabic
I. Hossen et al, 2019 [14]	Sentiment Analysis of Malay Social Media Text	Opinion Review	Text	Malay	Lexicon dictionary in Malay
A. Shrestha et al, 2019 [6]	Detecting depressed users in online forums	Depression Detection	Text	English	Detecting depressed users in online forums
Xinyu Wang et al, 2013 [53]	A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network	Depression Detection	Text	Chinese	Applied 3 kinds of classifiers over 180 users and improve the accuracy by 80%

Among the reviewed Articles above, the authors of [50] has adopted the Mixed method to make a comparison between Lexicon-Based Approach and Machine Learning Based Approach. In their comparison studies they have found that in both cases the effect of accuracy is almost the same. As they have used the parameter Text only.

In [7] researchers have performed an experiment by adding corpus as the second parameter beside Text. The specialty of adding corpus creates a new frame of understanding to the emotive statements. Unfortunately, the meaning of corpus changes based on the application of the language. Sometimes it contains a positive meaning with a negative representation of Language vice versa.

In Table 2 the articles related with Machine Learning are listed where single parameterized algorithm have been adopted except [54]. They have used Time and Frequency beside Text.

Although it's a good initiative in the field of Depression Detection, this algorithm fails to handle a large string sequence of Language. To reduce this problem Long Short Term Memory (LSTM) has been introduced with RNN.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Reviewed Articles on Machine Learning Based Approach

Author & Year	Title	Scope	Parameter	Language	Contribution
L. Rahman et al, 2018 [28]	Teenagers Sentiment Analysis from Social Network Data	Opinion Review	Text	English	Conceptual Framework
A. H. Uddin et al, 2019 [16]	Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique	Depression Detection	Text	Bangla	Depression Detection from Bengali tweets Using LSTM
A. U. Hassan et al, 2017 [49]	Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression	Depression Detection	Text	English	Find the depression level of a person
F. Cacheda et al, 2019 [55]	Early Detection of Depression : Social Network Analysis and Random Forest Techniques	Depression Detection	Text	English	Improved accuracy by 10%
M. M. Tadesse et al, 2019 [38]	Detection of depression-related posts in reddit social media forum	Depression Detection	Text	English	Evaluate a new model of depression detection by using NLP and ML
S. Arafin Mahtab et al, 2018 [23]	Sentiment Analysis on Bangladesh Cricket with Support Vector Machine	Opinion Review	Text	Bangla	Opinion Review on Cricket using SVM
M. H. Abd El-Jawad et al, 2018 [56]	Sentiment analysis of social media networks using machine learning	Opinion Review	Text	English	Compares the Machine Learning Classifiers and Introduced a new hybrid model
A. Kumar et al, 2019 [54]	Anxious Depression Prediction in Real-time Social Data	Depression Detection	Text, Time, Frequency	English	Investigate the time of post in the presence of anxiety and depression
Nafiz Al Asad et al, 2019 [57]	Social media content and sentiment analysis on consumer security breaches	Depression Detection	Text	English	Analyze the level of depression based on the quaternaries

After analyzing the previous works, we concluded that the previously conducted research works give us a good idea about the methods used in Sentiment Analysis. However, in the case of depression diagnosis, there are many uncertainties over the criteria that were used in the detection methods.

Based on the reviewed articles we have successfully extracted the research problems of the works, which gives us a big clue to investigate further. On the other hand, in previous studies the main attention has been drawn on finding the positivity and negativity of text. By dint of this point of view the statement has been marked as depressive or non-depressive. But for depression analysis it is very risky to state any statement as depressed without investigating any other symptoms of depression.

4. RESULT

From the above mentioned papers we have a clear idea about the social media that has been used vastly, about the terms of Sentiment Analysis and its levels. Facebook and Twitter have been chosen for sharing opinions worldwide. Here for social media data prediction Sentence level Sentiment analysis is used as it is the direct statement from people. The major approaches of Sentiment Analysis are Lexicon-based and Machine Learning. In terms of accuracy both shows equality in Figure 1.

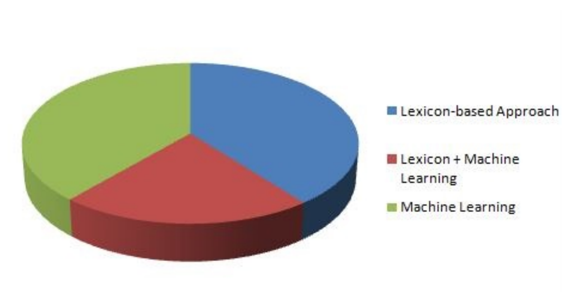

Figure 1: Accuracy of Methods

In Lexicon-based approach the whole analysis depends on the linguistic dictionary. The more precise the dictionary provides a more precise result. This approach is not intended to cover all facets of language, especially when it comes to slang, sarcasm and negation, owing to the complexities of natural languages [47]. It is not enough to use sentiment phrases. Any of the issues remain, such as some terms having varying

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

meanings depending on the application, some sentences consisting of words of sentiment but do not convey some emotions, and certain sentences without words of sentiment may still indicate opinion words [48]. The Lexicon-based system, however, has its own benefit, as it offers easy counting of positive and negative terms, versatility to match with various languages, and speed for maximum study [2].

The method of machine learning comes under the control of learning and allows training data to be analyzed. The mostly used classifier of Machine Learning is SVM, Naïve Bay's and RNN. These models depend of the training data. On the other hand sampling size of data is another concern, it fails to handle the random length of Facebook data and also for spelling mistakes Machine Learning requires a lot of training data sample [2]. In addition, the study of machine learning takes time as it takes hours in the complicated machine learning model, especially if training is needed [23]. After analyzing the result, we have a clear picture of the methods used in Depression Detection in Sentiment Analysis, which focuses the first objective of this study. Table 4 will describe the accuracy against domain individually.

Table 3: Accuracy Comparison of the studies

Author & Year	Domain	Accuracy
R. C. Dey et al, 2019 [24]	Facebook	91%
C. Dhaoui et al, 2017 [50]	Facebook + Twitter	45%
L. Rahman et al, 2018 [28]	Twitter	91%
X. Tao et al, 2016 [7]	Twitter	-
D. Doğru et al, 2018 [51]	Twitter	80%
S. Yuliyanti et al, 2017 [52]	Twitter	75%
M. Itani et al, 2017 [13]	Twitter	85%
I. Hossen et al, 2019 [14]	-	80%
A. Shrestha et al, 2019 [6]	Twitter	78%
A. H. Uddin et al, 2019 [16]	Twitter	82%
A. U. Hassan et al, 2017 [49]	Twitter	70%
F. Cacheda et al, 2019 [55]	Twitter	88%
M. M. Tadesse et al, 2019 [38]	Reddit	-
S. Arafin Mahtab et al, 2018 [23]	Online News	-
M. H. Abd El-Jawad et al, 2018 [56]	Twitter	70-85%
A. Kumar et al, 2019 [54]	Twitter	87%
Nafiz Al Asad et al, 2019 [57]	Twitter	74%
Xinyu Wang et al, 2013 [53]	Twitter	80%

5. DISCUSSION

Based on the reviewed papers, they mostly either demonstrate the Lexicon-based approach, Machine-learning approach or combine the methods. Additionally, the mentioned work focused on Twitter as their social media platform. Though Lexicon-based analysis is an advanced method to analyze the emotional aspect of text posted by the users, it has some drawbacks. This technique was designed for simple words used in the language. The complexity comes when it appears with slang, negation, and sarcasm [47]. The idea of using words to detect sentiment is not enough to conclude any decision strongly. Some problems occurred when dealing with this approach, such as some words contain different meanings in different applications, some sentences can express sentiment without using any sentiment words; on the other hand, some may not express any sentiment having sentiment words [48]. Especially in Bangla Language, very little attention was drawn to synthesize the Lexicon-base analysis. The major attention was paid to review the public opinion in Bangla Language. The accuracy of Machine Learning depends on the classifiers, so that it is important to follow up the gaps of the classifiers. The limitations of the Machine Learning classifier that has been used mostly are mentioned in section 2.7.2. To predict depression using language analysis is not enough [54]. In the text as written format, it is possible to find out the presence of positivity and the negativity of a text. Eventually it is possible to identify if the text contains any positive words or negative ones. We can relate these positive and negative words with our sentiment expressive behavior [25]. By the help of these words we can share the emotions that represent only our good or bad mood of mind [39]. Though it is helpful to express the happiness and sadness of our mind state, it is unable to clarify the depressive state of mind. More so because in depressive state we use the negative words to express our emotions which are similar to sadness [54]. So how can someone become so convinced of the depressive condition by only examining the negative presence in the text? It is not possible to state that anyone is suffering from depression, or sadness or any other issues. In a psychological research it has come to light that a depressive person has some other and specific symptoms and insomnia is one of them [4], [55]. Thus, it is important to analyze the time of posting along with the posted Text to ensure the presence of depressive symptoms. But the works on Bangla language lacked this attention on the time of post to predict depression. The problem formulation focuses on the gap of the existing work on Depression Detection. This discussion also raises some research questions:

- What are the methods to Detect Depression in Sentiment Analysis?
- What is the problem with the existing method of Depression Detection?

6. CONCLUSION

The conducted Literature Review demonstrates the pros and cons about Sentiment-Analysis and the level of it. It is also rich with the information of the approaches of Sentiment-Analysis, which is Lexicon-based approach and Machine Learning approach. The need of linguistic dictionary and its use in Bangla Language are clearly displayed in the reviewed papers. We have come to know about the classifiers of Machine Learning such as SVM, Navie Bayes, RNN and their limitations.

We have discussed about the depression and its relation of Social Media. To analyze depression from Social Media based on only text analysis is not enough. Symptoms of depression need to be included in the analysis. For example, the time of posting could indicate if a person is insomniac, a depressive symptom.

Moreover, this study will open a new scope of study in Bangla Language and also points out that we need more research on finding the other symptoms of depression. As in the terms of either clinical depression or general depression it is very important to identify other depressive symptoms like sleeping problems, anxiety, lose or gain of weight, suicidal tendency etc. This research will be a conductor for upcoming research on Bangla Language as the gaps and methods are clearly identified as well.

The future extensions of this work will add to the critical review of the algorithms and the Language filtering process to meet up the gaps of accuracy in the filtering as the present filtration techniques failed to identify the wrong representation on Language and to detect the weight of words based on their context.

REFERENCE

[1] B. R. Heravi, J. Mcleod, P. W. Eds, and D. Hutchison, *Transforming Digital Worlds*, vol. 10766. Springer International Publishing, 2018.

[2] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," in *Procedia Computer Science*, 2019, vol. 161, pp. 707–714, doi: 10.1016/j.procs.2019.11.174.

[3] C. Harper, "Diagnostic and statistical manual of mental disorders," *A Companion to Criminal Justice, Mental Health & Risk*. 2014, doi: 10.1097/00002093-198802020-00022.

[4] K. Andriessen, K. Krysinska, K. Kõlves, and N. Reavley, "Suicide Postvention Service Models and Guidelines 2014–2019: A Systematic Review," *Front. Psychol.*, vol. 10, no. November, 2019, doi: 10.3389/fpsyg.2019.02677.

[5] H. Almeida, A. Briand, and M. Meurs, "Detecting Early Risk of Depression from Social Media User-generated Content," 2017.

[6] A. Shrestha and F. Spezzano, "Detecting depressed users in online forums," *Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019*, pp. 945–951, 2019, doi: 10.1145/3341161.3343511.

[7] X. Tao, X. Zhou, J. Zhang, and J. Yong, "Sentiment analysis for depression detection on social networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10086 LNAI, pp. 807–810, 2016, doi: 10.1007/978-3-319-49586-6_59.

[8] G. S. Alexopoulos, "Vascular Disease, Depression, and Dementia," no. Ci, pp. 1178–1180, 2003.

[9] M. Rahman, S. Haque, and Z. Rahman, "Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique," *Int. J. Comput. Appl.*, vol. 176, no. 17, pp. 13–17, 2020, doi: 10.5120/ijca2020920119.

[10] A. H. Uddin, D. Bapery, and A. S. Mohammad Arif, "Depression Analysis of Bangla Social Media Data using Gated Recurrent Neural Network," *1st Int. Conf. Adv. Sci. Eng. Robot. Technol. 2019, ICASERT 2019*, vol. 2019, no. Icasert, pp. 1–6, 2019, doi: 10.1109/ICASERT.2019.8934455.

[11] M. Song, H. Park, and K. shik Shin,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

"Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 637–653, 2019, doi: 10.1016/j.ipm.2018.12.005.

[12] P. Sanguansat, "Paragraph2Vec-based sentiment analysis on social media for business in Thailand," *2016 8th Int. Conf. Knowl. Smart Technol. KST 2016*, pp. 175–178, 2016, doi: 10.1109/KST.2016.7440526.

[13] M. Itani, C. Roast, and S. Al-Khayatt, "Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media," *Procedia Comput. Sci.*, vol. 117, pp. 129–136, 2017, doi: 10.1016/j.procs.2017.10.101.

[14] I. Hossen, M. Goh, T. Connie, and A. Aris, "Sentiment Analysis of Malay Social Media Text," vol. 3, no. January 2019, pp. 241–251, 2018, doi: 10.1007/978-981-10-8276-4.

[15] H. Peng, E. Cambria, and A. Hussain, "A Review of Sentiment Analysis Research in Chinese Language," *Cognit. Comput.*, vol. 9, no. 4, pp. 423–435, 2017, doi: 10.1007/s12559-017-9470-8.

[16] A. H. Uddin, D. Bapery, and A. S. M. Arif, "Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique," *5th Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2019*, pp. 1–4, 2019, doi: 10.1109/IC4ME247184.2019.9036528.

[17] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/715730.

[18] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Comput. Sci. Rev.*, vol. 27, pp. 16–32, 2018, doi: 10.1016/j.cosrev.2017.10.002.

[19] L. Martin-Domingo, J. C. Martín, and G. Mandsberg, "Social media as a resource for sentiment analysis of Airport Service Quality (ASQ)," *J. Air Transp. Manag.*, vol. 78, no. January, pp. 106–115, 2019, doi: 10.1016/j.jairtraman.2019.01.004.

[20] C. Wong-Anuchit, C. Chantamit-o-pas, J. K. Schneider, and A. C. Mills, "Motivational Interviewing–Based Compliance/Adherence Therapy Interventions to Improve Psychiatric Symptoms of People With Severe Mental Illness: Meta-Analysis," *J. Am. Psychiatr. Nurses Assoc.*, vol. 25, no. 2, pp. 122–133, 2019, doi: 10.1177/1078390318761790.

[21] N. Suman, P. K. Gupta, and P. Sharma, "Analysis of stock price flow based on social media sentiments," *Proc. - 2017 Int. Conf. Next Gener. Comput. Inf. Syst. ICNGCIS 2017*, vol. i, pp. 64–68, 2018, doi: 10.1109/ICNGCIS.2017.34.

[22] N. Mishra and C. K. Jha, "Classification of Opinion Mining Techniques," *Int. J. Comput. Appl.*, vol. 56, no. 13, pp. 1–6, 2012, doi: 10.5120/8948-3122.

[23] S. Arafin Mahtab, N. Islam, and M. Mahfuzur Rahaman, "Sentiment Analysis on Bangladesh Cricket with Support Vector Machine," *2018 Int. Conf. Bangla Speech Lang. Process. ICBSLP 2018*, pp. 1–4, 2018, doi: 10.1109/ICBSLP.2018.8554585.

[24] R. C. Dey and O. Sarker, "Sentiment analysis on bengali text using lexicon based approach," *2019 22nd Int. Conf. Comput. Inf. Technol. ICCIT 2019*, pp. 18–20, 2019, doi: 10.1109/ICCIT48885.2019.9038250.

[25] S. Goswamil and J. Poray, "Human computer interaction for sentiment analysis and opinion mining: A review," *2016 Int. Conf. Comput. Electr. Commun. Eng. ICCECE 2016*, 2017, doi: 10.1109/ICCECE.2016.8009586.

[26] K. J. Giri, "International Journal of Advanced Research in Big Data - Overview and Challenges," no. July 2014, 2015.

[27] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017, doi: 10.1016/j.jbusres.2016.08.001.

[28] L. Rahman, G. Sarowar, and S. Kamal, "Teenagers Sentiment Analysis from Social Network Data," *Soc. Networks Sci. Des. Implementation, Secur. Challenges*, pp. 3–23, 2018, doi: 10.1007/978-3-319-90059-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

9_1.

[29] "• Malaysia: Facebook users 2018-2023 | Statista." https://www.statista.com/statistics/490484/number-of-malaysia-Facebook-users/ (accessed Dec. 01, 2020).

[30] "Singapore: number of Facebook users 2023 | Statista," *Statista Research Department*, 2020. https://www.statista.com/statistics/490492/number-of-singapore-Facebook-users/ (accessed Dec. 01, 2020).

[31] "• Thailand: number of Facebook users 2017-2025 | Statista." https://www.statista.com/statistics/490467/number-of-thailand-Facebook-users/ (accessed Dec. 01, 2020).

[32] "Facebook users in Bangladesh - March 2020 | NapoleonCat." https://napoleoncat.com/stats/Facebook-users-in-bangladesh/2020/03 (accessed Dec. 01, 2020).

[33] "Social Media users in Bangladesh - March 2020 | NapoleonCat." https://napoleoncat.com/stats/social-media-users-in-bangladesh/2020/03 (accessed Dec. 01, 2020).

[34] "Social Media Stats Bangladesh | StatCounter Global Stats." https://gs.statcounter.com/social-media-stats/all/bangladesh (accessed Dec. 01, 2020).

[35] S. A. El Rahman, F. A. Alotaibi, and W. A. Alshehri, "Sentiment Analysis of Twitter Data," *2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019*, 2019, doi: 10.1109/ICCISci.2019.8716464.

[36] K. Ramesh, M. Anto Bennet, T. Mari Arathy, V. Kavia, S. Preetha, and G. Nivedha, "Analysis of the attitude of social network analysis using r3 tool," *Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019*, no. Iccmc, pp. 998–1003, 2019, doi: 10.1109/ICCMC.2019.8819774.

[37] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in reddit social media forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.

[38] D. A. Baker and G. P. Algorta, "The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies," *Cyberpsychology, Behav. Soc. Netw.*, vol. 19, no. 11, pp. 638–648, 2016, doi: 10.1089/cyber.2016.0206.

[39] E. M. Seabrook, M. L. Kern, and N. S. Rickard, "Social Networking Sites, Depression, and Anxiety: A Systematic Review," *JMIR Ment. Heal.*, vol. 3, no. 4, p. e50, 2016, doi: 10.2196/mental.5842.

[40] M. P. Hamm *et al.*, "Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies," *JAMA Pediatr.*, vol. 169, no. 8, pp. 770–777, 2015, doi: 10.1001/jamapediatrics.2015.0944.

[41] K. Lup, L. Trub, and L. Rosenthal, "Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed," *Cyberpsychology, Behav. Soc. Netw.*, vol. 18, no. 5, pp. 247–252, 2015, doi: 10.1089/cyber.2014.0560.

[42] M. K. Ying, "Association between Use of Social Networking Media and Depression in Adolescents and Young Adults: A Systematic Review," 2016.

[43] E. Guedes *et al.*, "Internet Addiction and Excessive Social Networks Use: What About Facebook?," *Clin. Pract. Epidemiol. Ment. Heal.*, vol. 12, no. 1, pp. 43–48, 2016, doi: 10.2174/1745017901612010043.

[44] S. Wu, "The Relationship Between Social Media Use and Depression in College Students : a Scoping Review," 2019.

[45] M. A. Mamun, M. S. Hossain, and M. D. Griffiths, "Mental Health Problems and Associated Predictors Among Bangladeshi Students," *Int. J. Ment. Health Addict.*, 2019, doi: 10.1007/s11469-019-00144-8.

[46] M. T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, and K. H. Khan, "Sentiment analysis and the complex natural language," *Complex Adapt. Syst. Model.*, vol. 4, no. 1, 2016, doi: 10.1186/s40294-016-0016-9.

[47] M. T. Aziz, "Sentiment Analysis On Facebook Group Using Lexicon Based Approach," *2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol.*, pp. 1–4, 2016, doi: 10.1109/CEEICT.2016.7873080.

[48] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," *Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017*, vol. 2017-Decem, pp. 138–140, 2017, doi: 10.1109/ICTC.2017.8190959.

[49] C. Dhaoui, C. M. Webster, and L. P. Tan, "Social media sentiment analysis: lexicon versus machine learning," *J. Consum. Mark.*, vol. 34, no. 6, pp. 480–488, 2017, doi: 10.1108/JCM-03-2017-2141.

[50] Karamollaoğlu, D. Doğru, and U. & Yıldız, "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach," pp. 5–9, 2018.

[51] S. Yuliyanti, T. Djatna, and H. Sukoco, "Sentiment mining of community development program evaluation based on social media," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 4, pp. 1858–1864, 2017, doi: 10.12928/TELKOMNIKA.v15i4.4633.

[52] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, "Early Detection of Depression : Social Network Analysis and Random Forest Techniques Corresponding Author :," vol. 21, 2019, doi: 10.2196/12554.

[53] M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment analysis of social media networks using machine learning," *ICENCO 2018 - 14th Int. Comput. Eng. Conf. Secur. Smart Soc.*, pp. 174–176, 2019, doi: 10.1109/ICENCO.2018.8636124.

[54] A. Kumar, A. Sharma, and A. Arora, "Anxious Depression Prediction in Real-time Social Data," vol. 2019, pp. 1–7, 2019.

[55] J. Hao and H. Dai, "Social media content and sentiment analysis on consumer security breaches," *J. Financ. Crime*, vol. 23, no. 4, pp. 855–869, 2016, doi: 10.1108/JFC-01-2016-0001.

[56] J. Hao and H. Dai, "Social media content and sentiment analysis on consumer security br-eaches," J. Financ. Crime, vol. 23, no. 4, pp. 855–869, 2016, doi: 10.1108/JFC-01-2016-0001.

[57] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7867 LNAI, pp. 201–213, 2013, doi: 10.1007/978-3-642-40319-4_18.