Sentiment Analysis Techniques for Depression Detection from Micro-blogging Social Media Posts

Sayyed Usman Ahmed[1], Tameem Ahmad[2], Nesar Ahmad[3]

Department of Computer Engineering, Z. H. College of Engineering & Technology, Faculty of Engineering and Technology, Aligarh Muslim University, Aligarh, 202002, India

[1]syedusmanahmed@zhcet.ac.in, [2]tameemahmad@gmail.com, [3]nesar.ahmad@gmail.com,

Abstract

In the last decade and so, the use of social media is becoming a part of daily life, especially for the young generation; its' use is at their fingertips. This extensive use of Social Network Services (SNS) is generating a huge volume of data by sharing thoughts in form of text, images, audio, and videos. The activities done by individuals at these SNS suggest many things about their personalities. Medical health scientists are interested to analyze these activities to find out the state of mind and depression disorders in human beings. Sometimes this state of mental depression disorder (MDD) can lead to a big loss. A timely evaluation can suggest initiating preventive measures to avoid devastating consequences. To support such people who are suffering from depression disorder more congenial methods are needed which can analyze and recognize such depressive activities. In this work, various sentiment-based algorithms are applied to detect depression signals from social media posts. A comparative study is presented to evaluate the efficacy of various standard machine learning algorithms in detecting depression signals. We also applied some regression-based techniques to study the barrier losses in machine learning algorithms. It is found that the Support Vector Machine (SVM) Classifier outperformed the other counterparts and produced an average accuracy of 92 percent for classifying suicidal (depressed) tweets from non-suicidal (Non-depressed) tweets.

Keywords: *Depression Detection, Emotion Analysis, Sentiment Analysis, Opinion Mining, Social Network Analysis, Microblogging, Twitter, Mental Illness Detection, Natural Language Processing.*

I. Introduction

Social Networking Services (SNSs) also refed as social networking sites or social media have now become much more popular in the previous two decades and become part of the day-to-day activities of a human being. SNS users are using it for professional as well as social communication including for sharing their life achievements, news and events, feelings, views, reviews, commercials and others[1]. These day-to-day activities of sharing content at SNS is producing an enormous quantity of data since its inception around 2002[1]. Some popular SNS includes Orkut [2] (Google's SNS got shut down in 2014), Facebook[3], Instagram[4], Twitter[5], and LinkedIn[6]. Other content-sharing platforms with features of social networking include YouTube (for video), Flicker (for image), Slideshare (for presentation) ResearchGate (Research Article Sharing, discussion forum) are getting popularity in a specific type of communication and purpose. These SNSs are similar in terms of their nature of collaboration and sharing.

Due to advancements in Data Analysis, SNSs and their analysis have gained wide popularity among researchers in a very short span of time [7], [8]. The following (Figure 1) is the projection of increased research work on social media analysis indexed at Elsevier's Scopus.

The presented graph (Figure 1) depicts the increase in

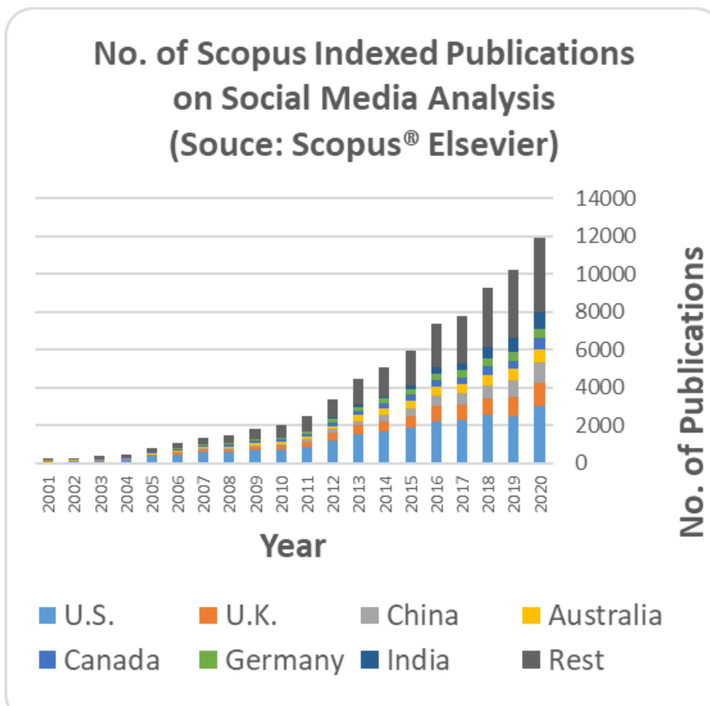


Figure 1: Publication Trends in Social Media Analytics

research work on data mining on social networks (Source: Scopus® Elsevier)[9]. The number of publications includes Journals Publications, Symposiums, proceedings (ACM, Springer, Elsevier, IEEE), Symposiums, Workshops etc. articles. This graph clearly shows an increase in the work and is about exponential. However, a large volume of data possesses difficulties in data analysis for researchers and data scientists. This suggests the need for a wider approach to incorporate cutting-edge technologies with Big Data to apply broader capabilities.

The study shows that an ample figure of young folks who did suicide shares a strong relationship with depression nearly at the instance of taking their own life voluntarily and intentionally[10]–[12]. In General, depression affects productivity that has side effects on the overall economy of the country. So, depression analysis, in today's time where young individuals widely use these Social Media Networks, is of prime importance.

At the same time, the huge amount of data is impossible to analyze manually. So, sentiment analysis is an efficient tool for automatic polarity generation of text as positive, negative and neutral. This can be done by classification techniques using different algorithms. These classification algorithms' performances are different for different types of tasks. This presented work is a comparative performance analysis of different chosen sentiment analysis algorithms (discussed in later sections III c).

II. Related Work

As the use of SNS is increasing, the researchers also attracted for analysis and explored the personalities and profiles for different purposes. Researchers are also attempted to explore the possibility of depression in the personality by examining their activities on social media websites [13]–[16]. However, the research in this field requires a higher level of effort to understand better the information shared through social websites and to develop a more robust system for real-time depression detection so that catastrophic results can be avoided timely.

Xinyu Wang et al.[17], in their paper, have studied depression detection on micro-blogs using data mining techniques. They believe that data generated through Social Network Systems (SNS) are extremely important for interdisciplinary studies like psychological and sociological studies[7], [18]–[20]. As their work was based on a Chinese dataset, they used ICTCLAS, a word segmentation system for the Chinese language [21], for sentence and word segmentation. In which each sentence S is broken down into sub-sentences Si and each sub-sentence is further broken into words and their respective Part-of-Speech (PoS) Wi{PSi}.

After successful segmentation, the polarity of the sentence is calculated based on man-made rules and a polarity calculation algorithm. The positive polarity signifies that a micro-blog has a positive sentiment and a negative polarity indicates a negative sentiment is depicted by the micro-blog. In case the polarity is computed as zero (0) then it means that the blog is objective. We can find out how positive or negative a sentiment is by its absolute value.

The depression detection model was based on Sina Micro-blog, a similar micro-blog as Twitter. They have analyzed the micro-blogs on the ten features like use of 1st person singular pronouns (I), use of 1st person plural pronouns (we), Positive and negative sentiments etc. For their experiment, they collected 6,013 micro-blogs and more than 50K sub-sentences were analyzed. To effectively classify the users into two categories: Normal Users and Depressed users, they have used the WeKa tool. Three different approaches were applied to this model for enhancing the reliability of the system namely, Naïve Bayes, Trees and Rules [22]. Their proposed model had a precision of about 80% with different classifiers. They applied a Binary Logistic Regression analysis using the SPSS tool for further simplification of the model and evaluation of the ten features identified earlier. The observation using this logistic regression was that the precision of the system declined by less than 5% however at the same time, this simplification had a good effect on the computing time and data collection.

Cavazos et al. in the paper [23] entitled "A content analysis of depression-related tweets" has worked to identify the clinical symptoms from the tweet texts manually by the experts of mental health research on a random sample of 2000 tweets. They coded them in five themes 1) disclosed depression 2) supportive message 3) depressive due to work-related pressures 4) in a situation to deal with depression and 5) disclosed self-harm or suicidal thoughts. Further, they added demographic details with profiles for further analyses on demographic characteristics using Demographics Pro. However, there is a need for an algorithmic approach to deal with and detect depression automatically, preferably on run-time to better deal with it.

Iram Fatima et al. in their research paper [24], analyzed the content generated by the users themselves from social communities instead of micro-blogs. They used these contents to predict and classify the degree of depression in the user-generated content. The sentiments and mood from user linguistics were studied using the mood tags and their respective valence value in the content. Their proposed model uses various machine learning approaches and some statistical tools to differentiate between a depressed and a non-depressed user. In their work, they identified depression at three levels: post, community and degree.

In studying depressive text covering a large set of data and avoiding misclassification of topics are two important challenges. Their work overcomes these two challenges by:

- Use of linguistic style and sentiment information from Linguistic Inquiry Word-Count [25] using feature extraction.
- Use of Hamilton Depression Rating Scale [26] values for classifying the degree of depression into three categories – mild, moderate and severe.

Data extraction is one of the initial and most important steps in studying social media data. The author has used LiveJournal for data collection using a crawler. The analysis of communities and extraction of features were done using LIWC Features and RELIEFE [27] respectively. For the classification of posts and communities, they have used an ensemble machine learning approach - Random Forest (RF). The level of depression in a post can be very well-identified by its mood tag and LiveJournal provide a rich set of 132 pre-defined mood tags. To map these mood tags with depression levels, the author used the ANEW lexicon [28], a tool that helps to manipulate the attributes of different tokens (words) in the corpus. To qualitatively measure the value of depression, they have used a fuzzy range of valence for the three categories of depression i.e., mild, moderate and severe. For training the algorithm and classifying the mood tags into class labels they have used a very well know probabilistic model known as Hidden Markov Model (HMM) which is based on Markov Chains. After successfully training the data, the author used the Baum-Welch algorithm to determine the depression level and their respective mood tags.

The proposed algorithm achieved an accuracy of 90% for correctly classifying a depressive post. Whereas the degree of depression had an accuracy of 92%. On a community level, the accuracy was the highest i.e., 95%.

III. Experiments

a. Proposal

Sentiment Analysis is extensively used in opinion mining. The opinion is mined using processing natural language and other techniques of computational linguistics and textual analysis[29]. In analyzing sentiments, we need to perform classification at various levels like documents or sentences. A collection of sentences forms a document [30]. The processing of sentiment includes some pre-defined steps: *Pre-processing of data, Feature extraction and selection, polarity calculation.*

Figure 2; shown below, describes these pre-defined steps as stated above.

Each of the steps may include many sub-steps like data processing includes tokenizing, removal of unwanted (stop) words, lemmatization etc.

The two basic methods for sentiment analysis are Lexicon-based and Machine Learning-based methods. In this presented work, we have tried to focus our work on the machine learning approach - supervised learning. The supervised learning approach used a labeled set of data for training. It uses different sets of data for training the model and prediction of the sentiments. Different types of supervised machine learning algorithms are available for sentiment analysis such as Naive Bayes, Support Vector Machine (SVM), Random Forest (RF), Linear regression, Artificial Neural Network (ANN), and Rule-based classifier.

This work has used - Linear regression, Support vector machine, random forest and Naive Bayes machine learning algorithms for processing social media data (tweets) and predicting sentiment polarity scores. These scores are indicative of the level of depression in those tweets.

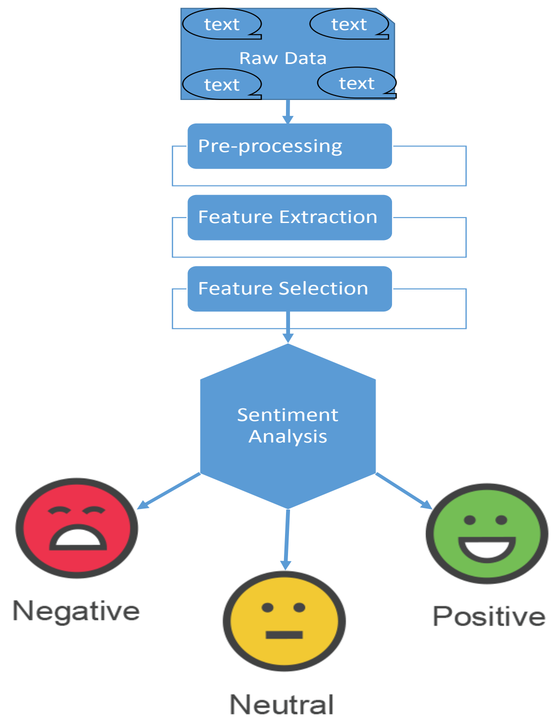


Figure 2: Data Pre-Processing Steps

The overall proposal of the work is presented below in figure 3. This figure represents the complete picture of the system from a data repository to the final classification of tweets into two This inclu ssing; vectorization, feature extraction and classification algorithm application viz. Naïve Bayes, Random Forest, SVM and Regression analysis.

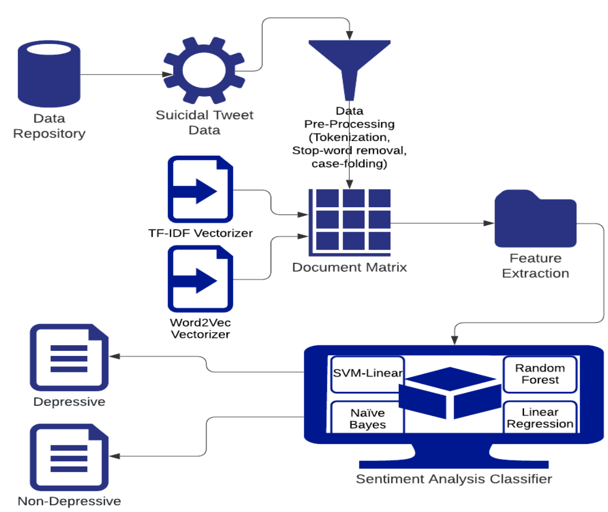


Figure 3: Overall System Architecture

b. Dataset

Selecting or constructing a dataset is crucial. The rationality of the dataset will contribute to the final results. Constructing a new dataset can be the most significant work with credibility. So, we have used an existing GitHub repository twitter-suicidal-intention-dataset by Laxmi Kant Tiwari (https://github.com/laxmimerit/twitter-suicidal-intention-dataset) [31], which has the labeled text data for the purpose of detecting the depression in tweets. The dataset contains about ten thousand (9119) cleaned Tweets with associated labels of classes. Label 1 indicated the class of depression (suicidal) and label 0 is for normal tweets (non-depressive). The following figure projects the data distribution in classes (Class 1 for depression and class 0 for non-depressive tweets). Further required processing of vectorization of tweets, before applying any classifier, is discussed in the later section. The dataset is saved into CSV files and provided to the system. This CSV file of the dataset contains roughly 50-50% of depressive and non-depressive tweets (Class 1 and 0). This 50-50 split dataset is adequate to train models for correctly classifying the tweets based on the content and making more precise predictions.

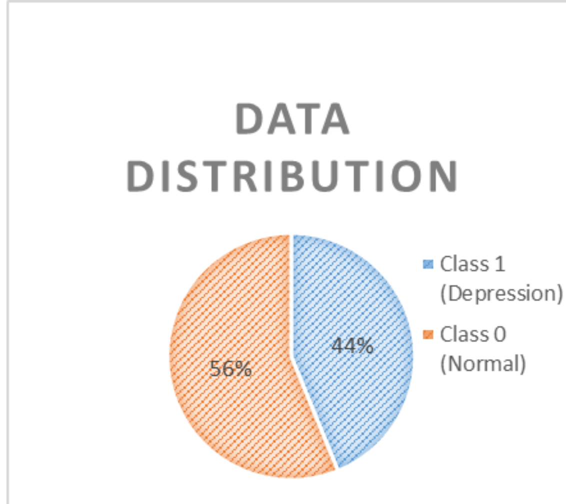


Figure 4: Dataset Distribution

Figure 5 shows the negative word clouds. These identified words (negative) are key players in depression detection. The presence of these negative words in the post are usually an indicator of negativity. This kind of posts can further be investigated for depression detection. Figure 5, Part (a) is a word cloud based on 140 selected negative words with their frequencies and part (b) word cloud is generated by 400 selected negative tweets from our dataset.

Figure 5:Negative Word Clouds (a) 140 Negative Keywords (b) extracted words from sample of 400 negative tweets [32].

c. Implementation Details

i. Vectorization

1. **TF-IDF:** The TF-IDF is a method to represent a document considering the words in the document. TF stands for Term-Frequency and IDF means Inverse-Document-Frequency. In Natural Language Processing a Document can be a single tweet or a collection of tweets (corpus).TF-IDF is a statistical measure used to determine the importance of a word in the complete document or corpus. Words in TF-IDF are assigned certain values called weights and this weight increases as the frequency of words increases in the document[33]–[35].

Mathematically it can be represented as:

$$TF(x) = \frac{No.of\ times\ x\ appears\ in\ the\ document}{Total\ No.of\ terms\ in\ the\ document}$$

$$IDF(x) = \frac{Total\ No.of\ documents}{No.of\ documents\ with\ term\ x\ in\ it}$$

$$TF - IDF\ (x) = TF(x) \times IDF(x)$$

2. **Word2Vec:** It is one of the advanced techniques for converting words in a document into their equivalent vector representation. It helps in efficient predictive modeling for word-embedding from raw data. The output of word2vec approach is a list of word vectors that represents a document/tweet in a corpus. Representing documents by vectors helps to understand the relationship (like similarity, difference etc.) of a word in the document[33].

There are two models on which this word2vec approach is based. One is **Continuous Bag-of-Words** (CBoW) and another is **Continuous Skip-Gram** (CSG). The difference between these two models is that CBoW is used to predict **a word** based on similar words in its surroundings whereas in CSG, we predict the surrounding words based on a word. The figure shown below represents these two approaches.

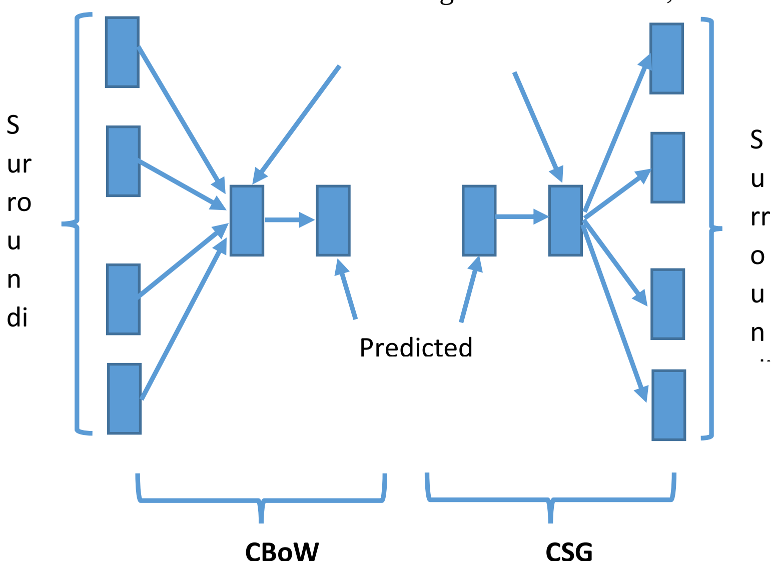


Figure 6: Word2Vec Representation

ii. **Naïve Bayes:** Humans are good at the classification of objects and we want to transfer this ability to machines using machine learning techniques[36]–[38]. Many examples of these intelligent classifications include recognizing the human gesture, sorting emails, assigning grades to HomeWorks etc. and many more.

Naïve Bayes is a kind of probabilistic classifier and methematically can be understood as for document d (say d), out of all the classes c ∈ C(c belongs to C) , this classifier returns the class ĉ which has the maximum posterior probability given the document given as below.

$$\hat{C} = \frac{P(d\,|\,c) * P(c)}{P(d)}$$

As P (d) is the same for all the classes, we can simplify the above equation by dropping P (d)

$$\hat{C} = P\ (d\ |\ c) * P\ (c)$$

P (d | c) = Likelihood of the document

P (c) = Prior Probability

iii. **Support Vector Machine (SVM):** Support Vector Machine is a supervised machine learning algorithm that distinguishes between two different data classes. This algorithm falls under the paradigm of supervised machine learning. It has a labeled training set for classification-related problems. SVM creates a hyperplane that depicts this binary classification of true/false or yes/no or negative/positive. For a given data, the algorithm with its excellent balancing capability can easily predict which class the data belongs.

An SVM-Linear classifier is the dot product of support vector X_i and a test row X^T. The following equation can depict the kernel function for an SVM-Linear:

$$K(X_i, X_j) = exp(-\frac{\|X_i - X_j\|^2}{2\sigma^2})$$

iv. **Random Forest**: A special type of classification algorithm creates multiple classifiers instead of using just one classifier. This type of classification is called Ensemble Classification. New data is correctly classified by taking into account the prediction of all the classifiers (also known as voting). We have used one such ensemble classifier in our sentiment analysis work i.e., Random Forest (RF) [39]. It is a tree-based classification and uses multiple trees in parallel for correctly classifying a value. A sample RF tree is shown below:

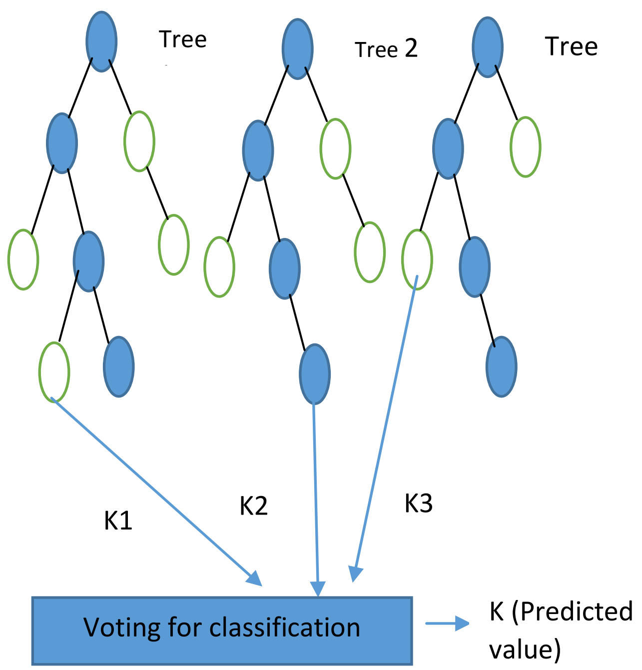


Figure 7: Random Forest Representation

v. **Logistic Regression**

This algorithm is admirably suited for discovering the link between features and some particular outcome. In Natural Language Processing (NLP) logistic regression is the baseline supervised machine learning algorithm for classification and also has a very close relationship with Neural Network. It classifies an observation into one of the two classes or one of many classes.

Logistic Regression is a discriminative classifier; they learn what input features are most useful to discriminate between different classes using **Sigmoid Function**.

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-(x.w+b)}}$$

Where,

w_i = weight, x_i = inputs and b_i =bias terms

vi. **Isotopic Regression and Platt Scaling**: Machine learning algorithms predict the probability of a correct class of data. Some algorithm like Logistic Regression has an in-built calibrated function like Sigmoid that helps calibrated values of probabilities. However, some algorithms require calibrations for better predictions of results, for example, support vector machines, random forest etc.

Platt scaling (Sigmoid) is used when the distortion in the predicted probabilities is of sigmoid shape. However, isotopic regression is applied in case the distortion is monotonic in nature.

We have applied both Platt Scaling and Isotopic regression for two machine learning algorithms – Naïve Bayes and Support Vector Machine (SVM) in order to calibrate the probabilities for these algorithms.

vii. **Evaluation**

Cross-validation of the generated result is conducted for feasibility by evaluating standard measures of accuracy, precision, recall, and F1 scores. Further recipient operating classification curves (ROCs) are also plotted.

The overall accuracy depends on the confusion matrix presented in table 1 below [40]:

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Table 1: Basic Structure of Confusion Matrix

IV. Results

The chosen classifiers are executed on the test dataset after preprocessing. As mentioned in previous sections, the chosen classifiers for comparative analysis are Naïve Bayes, SVM, Random Forest and Logistic Regression. Further, we have considered brier score loss for possible improvement. This inclusion differentiates our work from others and concentrates explicitly on improving with the possible decrease in Brier Loss value by reducing the mean square loss between predicted and actual outcomes. Moreover, we have attempted to improve the accuracy by applying Isotopic Regression and Platt Scaling (sigmoid) over Naïve Bayes and SVM. Table 1 depicts the brier loss for the four classifiers as mentioned earlier without isotopic and sigmoid.

Classifiers	Brier Loss	Avg. Precision	Avg. Recall	Avg. F1-score	Accuracy
Naïve Bayes	**0.213**	0.679	0.935	0.786	0.787
SVM	**0.128**	0.925	0.901	0.912	0.928
Random Forest	0.071	0.956	0825	0.885	0.887
Logistic Regression	0.070	0.939	0.840	0.887	0.910

Table 2: Results for different techniques without regression analysis

As we can see from Table 2, the loss for Naïve Bayes and SVM Classifiers are 21.3% and 12.8% which are pretty large in comparison to Random Forest and Logistic Regression classifiers. The figure shown below describes the calibration plot for these classifiers.

Figure 9: Calibration Plot after applying regression on Naïve Bayes

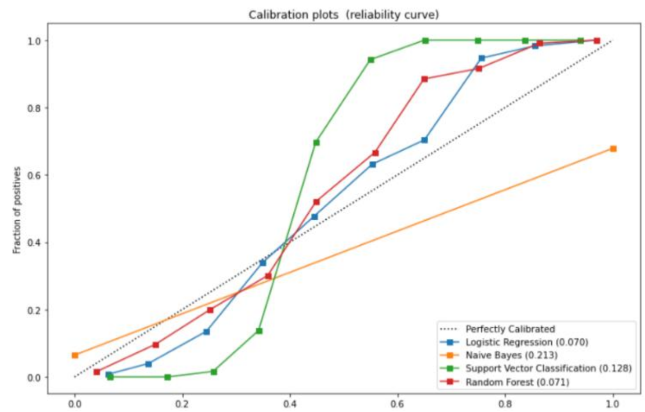


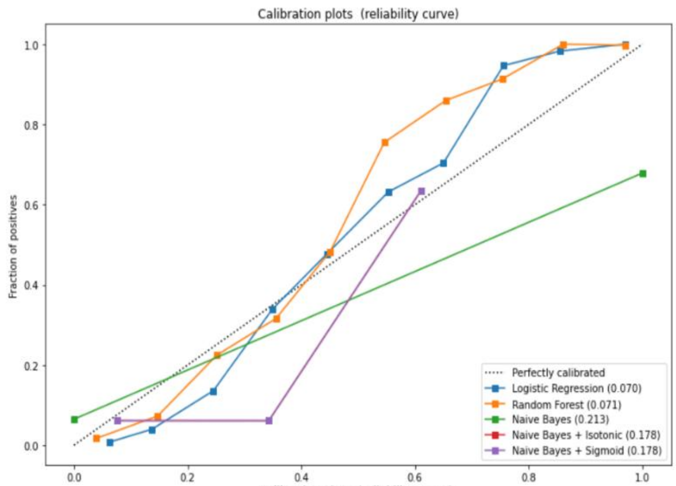
Figure 8: Calibration Plot (reliability curve) for classifiers without Isotopic or sigmoid

When we applied isotopic regression and sigmoid on Naïve Bayes and Support Vector Machine, we noticed a remarkable reduction in the brier loss. Table 3 reports this reduction.

Classifiers	Brier Loss	ΔLoss (Efficiency)	Avg. Precision	Avg. Recall	Avg. F1-score	Accuracy
Naïve Bayes	**0.213**	**0.035**	0.679	0.935	0.786	**0.787**
Naïve Bayes + Isotopic	**0.178**		0.635	0.945	0.760	**0.749**
Naïve Bayes + Sigmoid			0.635	0.945	0.760	**0.749**
SVM	**0.128**	**0.071**	0.925	0.901	0.912	**0.928**
SVM + Isotopic	**0.057**		0.927	0.885	0.906	**0.923**
SVM + Sigmoid			0.928	0.894	0.911	**0.927**
Random Forest	0.071		0.956	0825	0.885	0.887
Logistic Regression	0.070		0.939	0.840	0.887	0.910

Table 3: Results for different techniques with regression analysis

The figures below show the calibration plot after applying the regressions on Naïve Bayes and SVM.



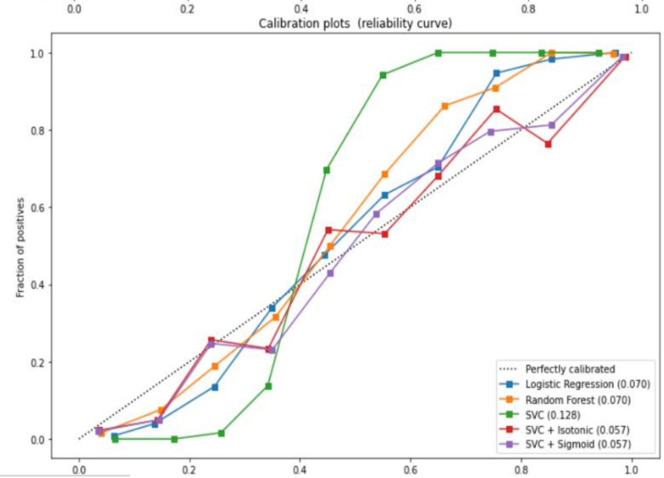


Figure 10: Calibration Plot after applying regression on SVM

By using regression on Naïve Bayes and SVM we achieved an efficiency of 3.5% and 7.1 % for Naïve Bayes and Support Vector Machine Classifiers respectively. Although we also observed there is a minor reduction in accuracy.

From among the classifiers we have used in this study, Support Vector Machine (SVM) classifier performed the best with an average accuracy of 92% for classifying the suicidal (depressed) tweets from non-suicidal (Non-depressed) tweets.

V. Conclusion & Future Scope

This work has studied depression recognition in social media data. To perform this detection automatically, using different classifiers, sentiment analysis, emotion analysis, data mining and text mining techniques are explored. Different classifiers have suggested that automatic depression detection is a capable indicator for mental health disorders and can recognize the abnormal behavior of the use. However, more congenial and reliable methods are needed to identify such mental disorders at run time for timely assistance and intervention to avoid any catastrophic results.

We examined state-of-the-art techniques for identifying Depression Disorders using Naïve Bayes, SVM, Random Forest and Logistic Regression and presented the comparative performance study. From the literature, we found that the text is the most important resource for identifying such disorders, followed by emojis and images audio and videos. Hence, in this study, the dataset was the text social media data in English. The study concluded that SVM Classifier outperformed the others with an average accuracy of 92% for classifying suicidal (depressed) tweets from non-suicidal (Non-depressed) tweets. Logistic Regression comes up next with 91% and then Random Forest with 88% and lastly, Naïve Bayes with 78% accuracy. We also applied regression analysis to study the barrier loss in these classifiers.

At a glance, the presented work is a comparative study based on standard machine learning algorithms. This is very much required to fill in the research gap in the study of depression detection. It will surely serve as useful pointers for other researchers working in this field. The challenge for this work is to handle satires and word sense disambiguation. Segregation of temporal events and their corresponding behaviors is another big challenge that can be dealt with by profiling and considering long-term events. Multilingual capability can be considered as the future scope of the work.

Health professionals and clinical advisors can use these results to streamline some methods for timely remedial or preventive action and assistance to overcome these situations and support and advice such people to help them to overcome.

VI. Acknowledgment

This work was supported by the Visvesvaraya Ph.D. Scheme for Electronics and IT fellowship of Ministry of Electronics and Information Technology (Meity), with awardee number, MEITY-PHD-2979, MEITY-PHD-3137, Government of India.

VII. References

[1] PEW RESEARCH CENTER, "Social Media Fact Sheet," 2019. [Online]. Available: https://www.pewresearch.org/internet/fact-sheet/social-media/. [Accessed: 06-Nov-2020].

[2] Orkut, "Orkut," *Google*. [Online]. Available: http://www.orkut.com/index.html. [Accessed: 06-Nov-2020].

[3] Facebook, "Facebook." [Online]. Available: https://www.facebook.com/facebook. [Accessed: 06-Nov-2020].

[4] Instagram, "Instagram." [Online]. Available: www.instagram.com/. [Accessed: 06-Nov-2020].

[5] Twitter, "Twitter." [Online]. Available: https://twitter.com/explore. [Accessed: 06-Nov-2020].

[6] LinkedIn, "LinkedIn." [Online]. Available: https://about.linkedin.com/?trk=homepage-basic_footer-about. [Accessed: 06-Nov-2020].

[7] V. Varshney, A. Varshney, T. Ahmad, and A. M. Khan, "Recognising personality traits using social media," in *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017*, 2018.

[8] O. Raghib, E. Sharma, T. Ahmad, and F. Alam, "Emotion analysis and speech signal processing," in *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017*, 2018.

[9] Elsevier B.V., "Scopus preview - Scopus - Welcome to Scopus," *Welcome to Scopus Preview*, 2004. [Online]. Available: https://www.scopus.com/. [Accessed: 31-Dec-2020].

[10] S. J. Cash and J. A. Bridge, "Epidemiology of youth suicide and suicidal behavior," *Current Opinion in Pediatrics*. 2009.

[11] T. Ahmad, S. U. Ahmed, S. O. Ali, and R. Khan, "Beginning with exploring the way for rumor free social networks," *J. Stat. Manag. Syst.*, vol. 23, no. 2, pp. 231–238, 2020.

[12] S. Tariq *et al.*, "A Novel Co-Training-Based Approach for the Classification of Mental Illnesses Using Social Media Posts," *IEEE Access*, 2019.

[13] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in reddit social media forum," *IEEE Access*, 2019.

[14] C. Lin *et al.*, "SenseMood: Depression detection on social media," in *ICMR 2020 - Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020.

[15] N. V. Babu and E. G. M. Kanaga, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," *SN Comput. Sci.*, 2022.

[16] C. Zachlod, O. Samuel, A. Ochsner, and S. Werthmüller, "Analytics of social media data – State of characteristics and application," *J. Bus. Res.*, 2022.

[17] 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," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

[18] T. Ahmad and N. Ahmad, "A Simple Guide to Implement Data Retrieval through Natural Language Database Query Interface (NLDQ)," 2020, pp. 37–41.

[19] T. Ahmad *et al.*, "An ISM approach for analyzing the enablers in India's makerspace platforms," in *2017 14th Learning and Technology, Conference*, 2017, pp. 14–22.

[20] M. S. Shaukat, M. Tanzeem, T. Ahmad, and N. Ahmad, "Semantic similarity–based descriptive answer evaluation," in *Web Semantics*, Elsevier, 2021, pp. 221–231.

[21] "Introduction to ICTCLAS." [Online]. Available: http://sewm.pku.edu.cn/QA/reference/ICTCLAS/FreeICTCLAS/English.html. [Accessed: 17-Apr-2021].

[22] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and TechniquesHan, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. San Francisco, CA, itd: Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-381479-1.00001-0*. 2012.

[23] P. A. Cavazos-Rehg *et al.*, "A content analysis of depression-related tweets," *Comput. Human Behav.*, 2016.

[24] I. Fatima, H. Mukhtar, H. F. Ahmad, and K. Rajpoot, "Analysis of user-generated content from online social communities to characterise and predict depression degree," *J. Inf. Sci.*, 2018.

[25] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," *Austin, TX Univ. Texas Austin*, 2015.

[26] R. M. Bagby, A. G. Ryder, D. R. Schuller, and M. B. Marshall, "The Hamilton Depression Rating Scale: Has the gold standard become a lead weight?," *American Journal of Psychiatry*. 2004.

[27] Y. Sun and D. Wu, "A RELIEF based feature extraction algorithm," in *Society for Industrial and Applied Mathematics - 8th SIAM International Conference on Data Mining 2008, Proceedings in Applied Mathematics 130*, 2008.

[28] M. M. Bradley and P. J. P. P. J. Lang, "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Technical Report C-2. University of Florida, Gainesville, FL.," *Psychology*, 2010.

[29] F. G. Contratres, S. N. Alves-Souza, L. V. L. Filgueiras, and L. S. DeSouza, "Sentiment analysis of social network data for cold-start relief in recommender systems," in *Advances in*

Intelligent Systems and Computing, 2018.

[30] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, 2012.

[31] L. K. Tiwari, "twitter-suicidal-intention-dataset," 2020. [Online]. Available: https://github.com/laxmimerit/twitter-suicidal-intention-dataset. [Accessed: 15-Jun-2021].

[32] T. Ahmad, S. U. Ahmed, and N. Ahmad, "Detection of Depression Signals from Social Media Data," in *Smart Connected World*, 2021.

[33] R. Bartusiak, Ł. Augustyniak, T. Kajdanowicz, P. Kazienko, and M. Piasecki, "WordNet2Vec: Corpora agnostic word vectorization method," *Neurocomputing*, 2019.

[34] J. Thompson *et al.*, "Relevant Word Order Vectorization for Improved Natural Language Processing in Electronic Health Records," *Sci. Rep.*, 2019.

[35] H. Hotta, M. Kittaka, and M. Hagiwara, "Word vectorization using relations among words for neural network," *IEEJ Trans. Electron. Inf. Syst.*, 2010.

[36] M. Lin and D. W. Cheung, "Automatic tagging web services using machine learning techniques," in *Proceedings - 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014*, 2014.

[37] S. Wang, J. Ren, and R. Bai, "A Regularized Attribute Weighting Framework for Naive Bayes," *IEEE Access*, 2020.

[38] H. C. Kim, J. H. Park, D. W. Kim, and J. Lee, "Multilabel naïve Bayes classification considering label dependence," *Pattern Recognit. Lett.*, 2020.

[39] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *J. Supercomput.*, 2021.

[40] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, 2020.