DOI: 10.21917/ijsc.2021.0327

A NOVEL APPROACH TO PREDICT THE LEVEL OF SUICIDAL IDEATION ON SOCIAL NETWORKS USING MACHINE AND ENSEMBLE LEARNING

Syed Tanzeel Rabani, Qamar Rayees Khan and Akib Mohi Ud din Khanday
Department of Computer Sciences, Baba Ghulam Shah Badshah University, India

Abstract

COVID-19 pandemic has taken millions of lives across the globe. Besides the death toll, the ongoing pandemic has a significant impact on mental health. Researchers contributed in this field revealed that the previous epidemics/ pandemics have a substantial relationship with the elevated rates of suicide. A growing apprehension is that there will be a spike in suicidal cases in the COVID-19 also that will be the next challenge for the world. Social Networking Sites (SNS) are becoming a new way for people to express their thoughts without worrying about the social stigma associated with mental illness. The various risk factors related to suicide like hopelessness, insomnia, anxiety and depression, if anticipated, can help in preventing suicide, thereby avoiding for being the potential victim of this mental disease. This paper focuses on the association between COVID-19 pandemics and suicidal ideation posts/tweets. A novel approach based on machine learning techniques is proposed for detecting and classifying suicidal posts/tweets into different levels of distress. This paper proposes a methodology through which relevant data related to suicidal ideation on social media is extracted, and a hybrid feature engineering mechanism is also proposed for obtaining the useful features from the dataset. The extracted features are then supplied to machine learning and ensemble learning models that automatically classify the risk of suicidal ideation into different levels (Multi-level classification) based upon their severity. Experiments reveal that F-measure is ranging from 0.69-0.98 with the best performance achieved through a Decision tree and Bagging approach. These findings emphasise and encourage the need for automatic classification of suicidal posts in the COVID-19 crisis and the possibility of building and using that machine learning model for immediate suicide risk screening.

Keywords:

Covid19, Suicidal Ideation, Machine Learning, Ensemble Learning, Multi-Level Classification

1. INTRODUCTION

The Covid-19 virus made its entry in late 2019 in Wuhan, China. The first confirmed case in India was reported on 30th January 2020. Presently, India has the highest number of confirmed cases in Asia and stands at third position after the United States and Brazil as per the global confirmed cases [1]. As on December 4, 2020, total worldwide figures of positive cases are 65.3 million, and 1.51 million deaths have been reported. In India, the confirmed positive cases are 9.57 million and more than I lakh and thirty nine thousand unfortunate deaths have occurred. Public health professionals have all their focus on testing and treating the patients who are infected with the virus but pay less attention to its psychiatric aspect. The COVID-19 has legitimately caused mental trouble in numerous people, with such impacts stemming not just from vulnerability about the future and disturbances to social help systems, yet in addition from critical financial difficulties, fears of physical security, and the experience of ailment and death among loved ones. Critically, vast numbers of these factors have been connected to severe mental pain bringing about self-destructive contemplations and suicidal thoughts. Studies have shown that uncertainty is looming large about the end to this pandemic situation that can take months and years before a vaccine will be available [2]. That will increase the psychiatric disorders, and suicide rates will upsurge [3]

Suicide is one of the leading public health concern worldwide consuming 8,00,000 lives per year, and on an average, a suicide occurs every forty seconds [4], [5]. Among the total global deaths due to suicide, almost 1,35,000 are from India alone [6]. World Health Organisation (WHO) and the American Foundation for Suicide Prevention (AFSP) identifies various risk factors that decrease the threshold for suicide. Some of the critical risk factors are anxiety, hopelessness, insomnia, schizophrenia, substance abuse, loss of family members, depression and previous suicidal attempts. Suicide is a common method considered to be a permanent solution to a temporary problem. In the early stages of the suicidal process, which starts with suicidal ideation, if the risk factors are understood, and the patient is referred for treatment, many lives can be saved. But the social stigma and a very fewer number psychiatrists and psychologists associated with the mental illness help prevent the treatment.

The research revealed that with the growing number of Social Networking Sites (SNS), People do feel comfortable to talk about their mental health freely on these sites without worrying about social stigma [7] [8]. Due to the unavailability of tests, lack of resources, and stigma associated with mental health, we believe that social media can provide the opportunity to analyse the behaviour of people through their posts and thus help the potential suicidal individuals.

Twitter is one of the famous SNS where people share their opinions, thoughts and feelings and is a searchable archive. Researchers have mined the twiiter for various issues like [9]. Twitter is functioning in every country except North Korea, Iran and China. Twitter posts can represent genuine feelings as people post without worrying about social stigma. It is estimated that 500 million tweets are tweeted every day, and the number of people who use Twitter for social networking is estimated to be 23% of the total number of people who are online [10]. Twitter recognised the presence of suicidal content and created a feature that helps to report such content. But this depends upon the interpretation of networked users and understanding the genuine risk is not possible by this such approach.

Reddit is another online discussion forum where people share their opinions. There are subreddits which allows people to share about a particular category. “SuicideWatch (SW)” is one of the subreddits where people share their potential suicide ideas. The Moderators are responsible for maintain the subreddit.

People experiencing suicidal ideation utilise various features in their language to depict about their distress related feelings which if anticipated ahead of time can help in preventing loss of valuable lives. The study conducted by us is a novel study that determines whether it is feasible to analyse the distress related tweets in Covid-19 by manually annotating the tweets into different levels of concern and then replicate the accuracy through machine classification.

2. RELATED WORK

The disaster caused by COVID19 pandemic is unimaginable. The social isolation, fear of losing job, anxiety and death of loved ones may lead to the accumulation of stress and finally to the suicide as seen in epidemics like Spanish Flu and Severe Acute Respiratory Syndrome (SARS) outbreak [11]. In our study, the problem of identifying people with suicidal ideation is treated as a text classification where stress-related tweets are collected and is used to train a classification model. Classifying the tweets into different levels of distress is still in its infancy. This problem has usually been dealt through psychiatric parameters where Questionaries' and clinical methods are used [12], [13]. Due to the data scarcity, researchers are moving towards the machine learning approaches to understand the language on social media as a viable option for understanding the suicidal ideation [14]-[16] Very little work has been done in the area of Natural Language Processing (NLP) and machine learning to understand the suicidal notes on social media. Jashinsky et al. [17] used a keyword-based approach to detect suicidal tweets. The collected geo-located tweets were matched with the random users in the United States, and a strong correlation was detected between the collected suicidal tweets and actual suicide rates. Bart Desmet et al. [5] used two machine learning algorithms viz. Support Vector Machine (SVM) and Naïve Bayes (NB) to automatically classify the suicidal posts from Dutch social media. The model optimisation was done through genetic algorithms that provided a good F-score. Bridianne O'Dea et al. [18] collected various keywords and phrases from suicide forums and extracted the relevant tweets using those keywords. The tweets were manually annotated and human accuracy was replicated through machine learning methods. The features provided to the machine learning model were unigrams and Bag of Words (BOW). Coppersmith et al. [19] analysed the suicidal tweets that were posted by the users before their suicidal attempt. After analysis, it was found that there was a spike in the sadness related tweets before the week of suicidal attempt. Thereafter there was an increase in anger tweets.

After critically analysing the literature regarding automatic classification of suicidal tweets, it was found that machine learning models were not performing up to the mark due to the data scarcity and traditional feature extraction mechanism. Moreover, the literature focussed of binary classification which doesn't give a clear idea about the level of distress. In our proposed methodology, we trained the multi-class machine learning model with rich features extracted through our proposed hybrid machine learning mechanism that resulted in high precision and recall value.

3. MOTIVATION/CONTRIBUTION

Our work makes novel contributions to the already existing literature in the following ways.

- A novel dataset of 10260 is prepared by extracting the tweets in 8 months and annotating them into different levels of distress by following the annotation scheme of data in consultation with psychiatrists and psychologists.
- A hybrid feature engineering technique is proposed that help in extracting the relevant features for better classification.
- Hidden knowledge is discovered from the data mining approach.
- Benchmarking is done on various machine learning and ensemble approaches to distinguish between the accuracy of learning models.

4. PROPOSED METHODOLOGY

The methodology proposed to train our machine learning and ensemble models for automatic classification is proposed by having an extensive discussion with mental health experts. The overall methodology consists of five steps (i) corpus collection and human annotation (iii) pre-processing (iv) feature engineering (v) classification (vi) evaluation and validation.

The graphical representation of the methodology is shown in Fig.1.

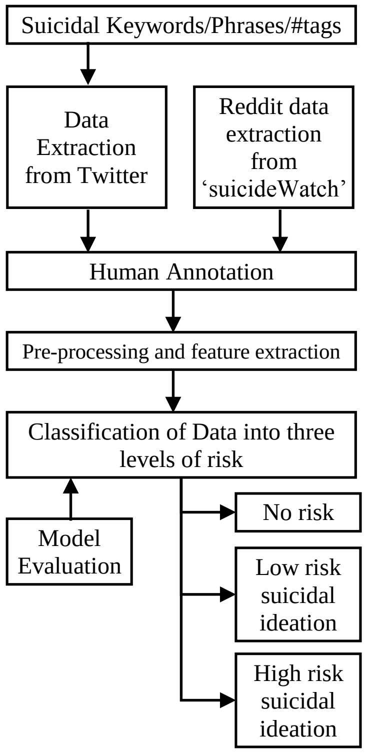

Fig.1. Proposed Methodology for Multi-level suicide risk classification

4.1 CORPUS COLLECTION AND HUMAN ANNOTATION

4.1.1 Twitter Data Extraction:

The Application Programming Interface (API) of Twitter is used to collect the data using various keywords and phrases like "#Covid19Stress," "CovidAnxiety," "Corona19SleepDisorder," "Covid19Insomnia," "#depression," "#suicide." "#killmyself," "WantToEndMylife," "WantToDie" etc. A total of 22567 tweets are collected using keyword filtering technique that also contains sarcastic tweets and tweets related to advertisement also. Out of the total extracted tweets, 7582 tweets are annotated according to the annotation rules.

4.1.2 Reddit Dataset:

Reddit is an online discussion forum where people share their opinions. There are subreddits which allows people to share about a particular category. “SuicideWatch (SW)” is one of the subreddits where people share their potential suicide ideas. We used the Reddit interface to collect the 2678 suicidal posts and some of the posts were borrowed from [15]. The posts are annotated according to the annotation rules and subsequently merged into the already extracted twitter dataset that makes a large dataset of about 10260 Posts.

On trying to understand the dataset in its first place, we try to comprehend the frequent words in the dataset that are used by suicidal individuals and the length of those post that people use for different levels of distress. Word cloud in Fig.2 revealed that the words “want,” “kill” “feel”, “die,” “I’m” “corona stress,” “depression” are used in abundance. For example, People with high-risk use the words like “wish,” “want” in a post as “I wish to kill myself right now”. People with low risk use the words like “feel” in their post as “I feel I am drowning”. The Fig.3 revealed that the higher risk individuals use more length in their post to discuss their suicidal ideation than lower-risk individuals.

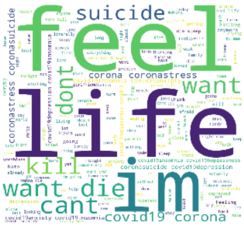

Fig.2. Word cloud of the extracted Data

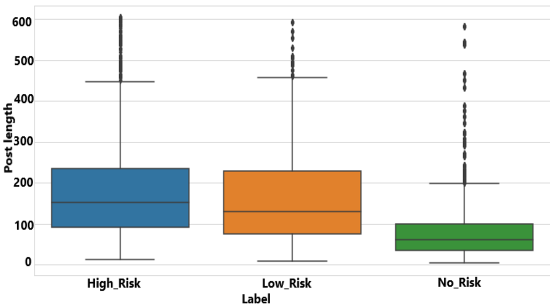

Fig.3. Tweet Length vs. Risk Category

4.1.3 Human Annotation:

Human annotation is an important step for training the supervised machine learning model. The annotation is performed in consultation with the mental health experts due to the nature of the problem. The Tweets/ Posts are labelled according to their severity in the three classes. According to the annotation rules, Tweets/posts with high risk are those that talk about killing oneself, having a suicidal idea or wishing about it. Low-risk tweets are those that talk about depression, anxiety, hopelessness or any indirect reference to the suicide. No risk tweets/posts are those that sarcastically speak about suicide, suicide campaigns or second person suicidal references.

4.2 DATA PRE-PROCESSING

As social media data is very noisy, it is necessary to reduce the noise and improve the data quality. The established methods are used to pre-process the data that will help in improving the classification results [20]. The tweets/posts are tokenised, stop words removed, and lemmatisation is performed to get the root word.

4.3 FEATURE ENGINEERING

A hybrid feature engineering technique is used to extract the rich features that are then supplied to the machine learning model for classification. We improved the technique of [15] to extract the features using Term Frequency Inverse Document Frequency (TFIDF), (BOW) , Post length and Topic Modeling. After that, we use the Information gain algorithm, along with ranker search method to find the most significant features. The threshold for discarding the less significant attributes is set to 0.0020. The maximum limit of the feature set is set to 424. We use BOW that consists of words as features up to trigrams to preserve the information. BOW don’t take the context of posts in consideration, so we also use TFIDF to extract the relevant features. TFIDF is a technique that assigns the score to the particular feature based upon its presence in the particular document and whole corpus. In the context of our problem TFIDF is described by the following equation.

$$tfidf(f,p,D)=tf(f,p)\times idf(p,D) \qquad (1)$$

$$idf(f,D)=\log\frac{|D|}{1\left|\{p\in D: f\in p\}\right|} \qquad (2)$$

where f denotes the word as a feature, p denotes the post from which features is extracted, and D represents the document space (set of all posts and tweets)

The post length was also used as a feature as the length of suicidal and non-suicidal text significantly varies as revealed by Fig.3. The statistical features are calculated based upon the number of words and characters in the sentence and number of sentences in a post. Part of speech (POS) also distinguishes the text by considering the similar kind of grammar used in the text body. Various POS tags that are usually used are nouns, adverbs, articles and pronouns.

Topic features are also used to extract the 10 different topics from the suicidal posts using Latent Dirichlet Allocation (LDA) [21]. The Overall picture of Feature Engineering is shown below in Fig.4.

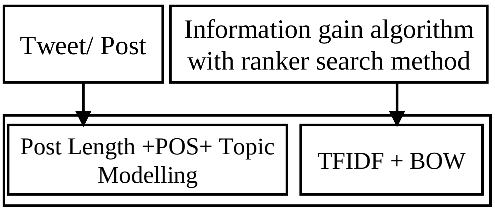

Fig.4. Feature Engineering for extracting the relevant features

4.4 CLASSIFICATION AND VALIDATION

The detection of suicidal ideation in our research work is treated as a multi-class machine learning problem. The dataset we refined consists of the only two columns, the title of text and the label. We formulated our problem by getting the motivation and ideas from [15]. We have a corpus $\{p_i, l_i\}_i^n$ consisting of a set of posts/tweets $\{p_i\}_i^n$ and labels $\{l_i\}_i^n$, the training is provided in a such a way that model learns from the data and the corresponding labels. The supervisory function that guides the model is as follows

$$l_i = Fun(p_i) \tag{3}$$

When s post p_i represents the high risk of suicide, l_i=2. When post p_i represents the low risk, l_i=1 and l_i=0 in case of p_i representing the no risk of suicide.

We experimented out classification task using four Machine learning Algorithms as Decision Tree, Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression. Three Ensemble methods as, Bagging, AdaBoost and Stochastic Gradient Boosting are also used to classify the suicidal data into different levels of risk. We evaluated the model using Confusion metrics and validated the model using 10 cross-validation technique as discussed below in Results and Discussion section.

5. RESULTS AND DISCUSSION

We used Twitter API to collect a total of 22567 suicidal tweets using different keywords and phrases [18] [22] borrowed from various sites and previous research papers that discuss suicidality. We also used the social media forum Reddit to extract the 2676 posts using its subreddit 'SuicideWatch'. After pre-processing the posts/tweets using the various established methods [20], the tweets are annotated into various levels of risk in consultation with mental health experts. The proposed hybrid feature extracting technique combining TFIDF, BOW, Length of Tweets, POS and Topic features is applied to extract the features. Most relevant features are then calculated using Information gain algorithm with ranker searcher method. A total of 424 significant features are then supplied to machine learning and ensemble methods for classification. Dataset is split into 70:30 ratio, where 70% of the data is used for training the model, and 30% is used to test the model. Python language is used to implement the machine and ensemble learning algorithms. The different packages that are used are NLTK, Sckitlearn, pandas etc. The Table.1 shows in detail the performance of various applied learning algorithms.

The performance and accuracy of the classification model is evaluated through various metrics like Precision, Recall and F-measure. Among all the classifiers, Decision tree and Bagging approach were best performers with Decision tree having F-measure ranging from 0.96-0.98 and Bagging has also the same F-measure ranging from 0.95-0.98. We selected F-measure to show the comparison of the accuracy of various algorithms below in Fig.5. The reason for selecting this measure is its ability to balance the influence of Precision and Recall.

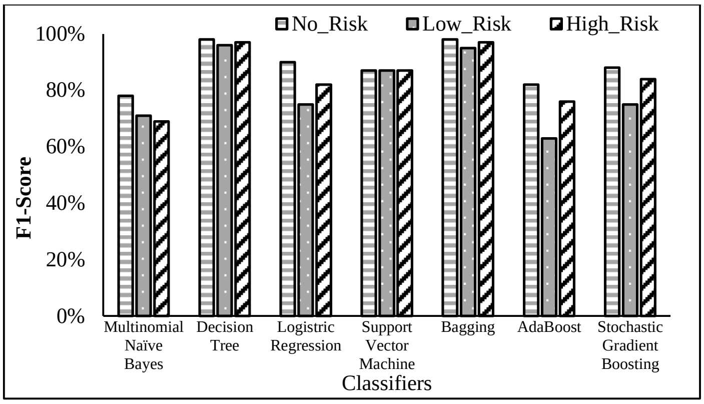

Fig.5. Comparison of Various machine learning and ensemble methods

The confusion metrics of the top most performing machine learning and ensemble methods are shown from Fig.6-7.

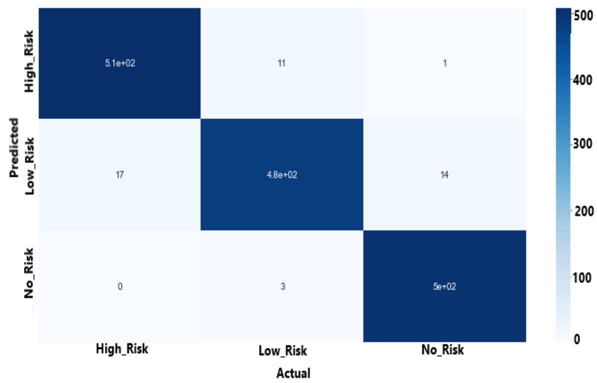

Fig.6. Confusion metric of Decision tree algorithm

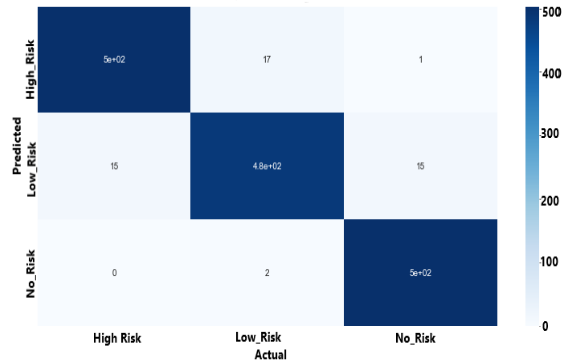

Fig.7. Confusion metric of Bagging

As per our insights in the literature, we found that our's is the novel work in COVID19 that provides the better identification and classification of suicidal tweets. This research work also provides a better feature processing technique for the identification of various features. Further in our dataset, there is no issue of under skewness. During the training and testing our model, there was neither underfitting nor overfitting issues. For validating the model, we performed the 10 cross-validations on our dataset.

6. CONCLUSION AND FUTURE WORK

With the exponential growth in social networking sites (SNS) and people feeling comfortable to post their thoughts on SNS, there is a dire need to channelise this medium for suicide prevention. We claim that most of the research in suicide detection and prevention has been done through face to face settings using clinical methods. The stigma is a significant barrier which prevents people from consulting the psychiatrist or psychologist. The less availability of psychiatrists is another hurdle for preventing potential suicide individuals. Further, Covid-19 amplified the risk of suicide by increasing the risk factor of social isolation and stress of going to consult the mental health expert due to the fear of contracting the virus. There arise the need for an alternate method which can analyse the potential suicide individuals without worrying about the social stigma and fear of Covid-19.

In this article, we formulated the problem of detecting suicidal ideation as a text classification problem. The related data is collected from Twitter and Reddit, pre-processed, and the proposed feature extraction technique is applied for the extraction of relevant features. The features are then supplied to machine learning and ensemble methods for classifying the suicidal data into different levels of distress based on their severity.

The various Algorithms that were applied are Decision Tree, Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression, Bagging, AdaBoost and Stochastic Gradient Boosting. Among all the machine learning algorithms and ensemble methods, Decision tree and Bagging outperformed all algorithms. The replication of machine accuracy to the extent of 99% indicated that machine learning methods could come rescue in solving this emerging mental health issue. This research work provides a base for researchers to solve this problem considering the other aspects. Some of the Future directions on which we will work are listed below

- The communication and connectivity between suicidal users will be identified.
- Questionnaires will be used to capture the inputs and will be used in our machine learning algorithms
- More refined categories like sadness, fear etc. will be used to classifying the data for better analysis

Deep learning algorithms will be investigated for better accuracy.

REFERENCES

[1] COVID-19 Pandemic in India - Wikipedia, Available at: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India, Accessed at 2020.

[2] M. Cascella, M. Rajnik, A. Cuomo, S.C. Dulebohn and R. Di Napoli, "Features, Evaluation and Treatment Coronavirus", Available at: ncbi.nlm.nih.gov/books/NBK554776/, Accessed at 2020.

[3] F. Ornell, J.B. Schuch, A.O. Sordi and F.H.P. Kessler, "Pandemic Fear and COVID-19: Mental Health Burden and Strategies", *Brazilian Journal of Psychiatry*, Vol. 42, No. 3, pp. 232-235, 2020.

[4] L. Sher and J.J. Peters, "The Impact of the COVID-19 Pandemic on Suicide Rates", *QJM: An International Journal of Medicine*, Vol. 113, No. 10, pp. 707-712, 2020.

[5] B. Desmet and V. Hoste, "Online Suicide Prevention through Optimised Text Classification", *Information Science*, Vol. 439-440, pp. 61-78, 2018.

[6] World Health Organization, "Suicide", Available: https://www.who.int/news-room/fact-sheets/detail/suicide, Accessed at 2020.

[7] V.K. Kailasam and E. Samuels, "Can Social Media help Mental Health Practitioners Prevent Suicides?", *Current Psychiatry Research and Reviews*, Vol. 14, No. 2, pp. 37-39, 2015.

[8] C. Mcclellan, M.M. Ali, R. Mutter, L. Kroutil and J. Landwehr, "Using Social Media to Monitor Mental Health Discussions-Evidence from Twitter", *Journal of the American Medical Informatics Association*, Vol. 24, No. 3, pp. 496-502, 2017.

[9] A.M.U.D. Khanday, Q.R. Khan and S.T. Rabani, "Identifying Propaganda from Online Social Networks During COVID-19 using Machine Learning Techniques", *International Journal of Information Technology*, Vol. 13, No. 1, pp. 115-122, 2020.

[10] Statista, "Twitter: Most Users by Country Statista", Available at: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. Accessed at 2020].

[11] J.A. Lieberman, "Meeting the Mental Health Challenge of the COVID-19 Pandemic Psychiatric Times", Available at: https://www.psychiatrictimes.com/view/meeting-mental-health-challenge-covid-19-pandemic, Accessed at 2020.

[12] M.K.M.T. Shioiri, A. Nishimura, K. Akazawa, R. Abe, H. Nushida and Y. Ueno, "Incidence of Note-Leaving Remains Constant Despite Increasing Suicide Rates", *Psychiatry and Clinical Neuroscience*, Vol. 59, No. 2, pp. 226-228, 2005.

[13] D. Sikander, "Predicting Risk of Suicide using Resting State Heart Rate", *Proceedings of Asia-Pacific Annual Summit and Conference on Signal and Information Processing Association*, pp. 1-4, 2016.

[14] E. Kiciman, M. Kumar, M. De Choudhury, G. Coppersmith and M. Dredze, "Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media", *Proceedings of International Conference on Human Factor Computer Systems*, pp. 2098-2110, 2016.

[15] S. Ji, C.P. Yu, S.F. Fung, S. Pan and G. Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content", *Complexity*, Vol. 2018, pp. 1-10, 2018.

[16] S.T. Rabani, Q.R. Khan and A.M.U.D. Khanday, "Detection of Suicidal Ideation on Twitter using Machine Learning and Ensemble Approaches", *Baghdad Science Journal*, Vol. 17, No. 4, pp. 1328-1339, 2020.

[17] J. Jashinsky, "Tracking Suicide Risk Factors through Twitter in the US", *Crisis*, Vol. 35, No. 1, pp. 51-59, 2014.

[18] B. O'Dea, S. Wan, P.J. Batterham, A.L. Calear, C. Paris and H. Christensen, "Detecting Suicidality on Twitter", *Internet Interventions*, Vol. 2, No. 2, pp. 183-188, 2015.

[19] G. Coppersmith, K. Ngo, R. Leary and A. Wood, "Exploratory Analysis of Social Media Prior to a Suicide Attempt", *Proceedings of International Conference on*

Computer Lingusitics Clinical Psychology, pp. 106-117, 2016.

[20] N. Anand, D. Goyal and T. Kumar, "Analyzing and Preprocessing the Twitter Data for Opinion Mining", *Proceedings of International Conference on Networks and Systems*, pp. 213-221, 2018.

[21] D.M. Blei, A.Y. Ng and M.I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 99-1033, 2013.

[22] American Foundation for Suicide Prevention, "Home-AFSP", Available: https://afsp.org/, Accessed at 2019.