IJCAT - International Journal of Computing and Technology, Volume 7, Issue 4, April 2020
ISSN (Online) : 2348-6090
www.IJCAT.org
Impact Factor: 0.835

Social Media Depression Monitoring Model Using Sentiment Analysis

[1] Ebunoluwa Okediran; [2] Jonah Joshua; [3]Afolashade Kuyoro; [4]Abolaji Akanbi

[1] Computer Science Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

[2] Computer Science Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

[3] Computer Science Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

[4] Computer Science Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State, Nigeria.

Abstract - The proliferation of mobile technology, with the privacy and ubiquity that it offers often presents social media pseudo-confidant for lonely and depressed individuals. Social media continues to play an active part in contemporary human life, according to statistics about 40% of the global population are currently active on social media. According to the World Health Organization, at least one in every twenty-five people living as at October 2019 are depressed. Several authors and researchers have developed and proposed methods to predict depression through sentiments in social media posts. Twitter have been widely investigated although other social networks have also been discoursed such as Sina (a Chinese micro blog) and Myspace. It is in light of this that this research seeks to gather social media dataset that is applicable for non-clinical early prediction of the mental health status of users towards strengthening business intelligence. Hence, this work is focusing on mining comments from twitter as a mitigation technique for depression.

Keywords - *Classification Algorithms, Depression, Machine Learning (ML), Sentiment Analysis (SA), Social Media.*

1. Introduction

Depression is the foremost reason of suicide globally, cutting across a wide demographic composition that spans gender, age, and occupation. Three and four-tenths of the global human population in 2017 were living with depression [1]. According to the World Health Organization, at least one in every twenty-five people living as at October 2019 are depressed [2]. This population results in suicidal deaths every forty seconds. The economic implications of depression, and the often-resulting suicidal consequences could portend serious implications for organizations; even as insurance businesses are also beginning to awaken to the reality that payment of death benefits are deserving even for cases when deaths are the result of suicide [3]. Also, the impact of this reality on the manpower and human capital resources of the global economy is a cause for concern, considering the fact that suicide is the second leading reason for death in adolescents and youthful grown-ups between the ages of 15-29 [3]. Furthermore, with the rising trends of workplace suicides, progressive have started to focus on the mental health of their employees as part of their business intelligence processes and practices [4]. The mandates of corporate social responsibility reasonably expect that responsible organizations take care of the mental health of the employees that help keep them running. The problem, however, is that mental health datasets are rare to come across, and a large percentage of those available are clinical datasets. Clinical datasets are a staple resource for most health and medical research. Clinical data is either collected during the course of on-going patient care or as a part of a formal clinical trial program. These datasets usually require clearance procedures that are often long and arduous before it can be obtained; and when these datasets are provided, they are ethically limited in their scope and extent of usage, since they are essentially confidential patient records. Social Media Monitoring powered by Artificial Intelligence (AI) and psychological innovations can give significantly progressively amazing knowledge from the data accumulated from social media. The act of social media data mining gathers and processes unstructured data, for example, posts, remarks, tweets, pictures shared on network platforms like facebook, instagram, myspace, twitter [5].

Several authors and researchers have developed and proposed methods to predict depression through sentiments in social media posts. Some of the proposed methods includes Crowdsourcing methodology for individuals diagnosed with clinical depression, utilization of an electronic survey for estimating the level of depression, ruminating behavior of depressed individuals, probabilistic model to determine depressive tweets, use of diverse machine learning algorithms for example: Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM). Linguistic Inquiry and Word Count library (LIWC) was additionally utilized by certain creators to accomplish text analysis and decide its extremity. In a similar setting, Sentiment Analysis utilizing LIWC was additionally put in to make a depression recognition model for Chinese micro-blogs. However, they have not been able to consider monitoring user's tweets on twitter assigning recommendations or treatments, also, aware warnings to a designated person as a follow up when depression is identified in an individual. It is in light of this that this study seeks to develop a depression monitoring model using Sentiment Analysis.

2. Literature Review

[6] Applied logistic regression with similar goal to characterize Myspace after death posts with (or not) emotional pain in this circumstance. Authors utilized the Linguistic Inquire and Word Count library (LIWC) to analyze words and gain extremity.

[7] To build a depression identification framework for Chinese micro-blogs, Sentiment Analysis (SA) utilizing LIWC was implemented in the same context. The model accumulated 10 highlights from the text (once more, the extremity of the text yet additionally other inventive highlights, for example, the utilization of emojis). The underlying 10 highlights were eventually decreased to 5, with a comparing debasement in the outcomes yet elevated enhancement in regards to estimation time and measure of information required.

[8] Propounded a probabilistic model to figure out which tweets could demonstrate depression. They remove both post-based (for example positive or negative impact, linguistic style) and user-based features (for example number of followers/followees, number of posting). Subsequent to utilizing dimensionality decrease Principal Component Analysis (PCA) and utilizing SVM they got 74% precision.

[9] Authors employed the utilization of an online survey for estimating the magnitude of Japanese twitter user.

[10] likewise analyzed Twitter with a similar reason. For this situation, they utilize a bag-of-words (sums the word appearance frequencies to evaluate the content of a tweet) to vectorize the tweets. At that point, they utilized diverse Machine Learning (ML) algorithms (LR, NB and SVM) and acquired around 80% exactness in identifying depression disorder.

[11] Developed a prototype targeting at signifying the mechanism of the method and probable public impact that might be conveyed. The commitments incorporate knowledge base of depressive sentiment and an algorithm to break down written information for depression recognition.

[12] Cause us to notice Reddit. Authors made and disseminated a dataset together with Reddit messages from various users, identifying some of the users with depression. The dataset was labeled demonstrating if the users were in danger of depression. The authors noticed generally the utilized classification measures (recall, F-measure, precision) are time-uninformed, which does not remunerate rapid alarms. In this manner, they proposed another measurement for early recognition that penalizes the delay in identifying positive cases. They likewise revealed the aftereffects of some fundamental baselines to distinguish rapid indications of depression. These baselines are very basic and the features removed from the text are simply founded on the tf-idf vectorization.

[13] The objective of the study introduced a monitoring solution for clients with potential mental upsets, explicitly stress. To accomplish the objective, a sentiment analysis metric to determine the person's mood is modelled and implemented. The metric is based on Portuguese language word-dictionary with enhancement of user's profile data, for example, age and gender.

[14] Proposed state of art early identification of depression via social media by methods for considering ground-breaking procedures that are explicit for rapid identification, mining text features in a more precise route than the present methodologies dependent on tf-idf vectorization, integrating SA, investigating the behavior of standard and current ML methods to more readily forecast positive cases.
Unlike existing and proposed models, this study uses the classification algorithm, Depressive Sentiment Vocabulary (DSV) and Valence Aware Dictionary for Sentiment Reasoning (VADER) model to determine the sentiment intensity which helps analyze the magnitude of depression.

3. Methodology

3.1 Existing System

The architecture of proposed system is illustrated in Figure 1: The system will first collect data (e.g., tweets posted by social network players) from social media. The textual data will then be processed and analyzed by the system communicating with a knowledge base storing depressive sentiment vocabulary. The analysis result, for instance, any issues regarding potential depression will then be alerted to the monitor (e.g., social workers, the parents of children) via a friendly user interface.

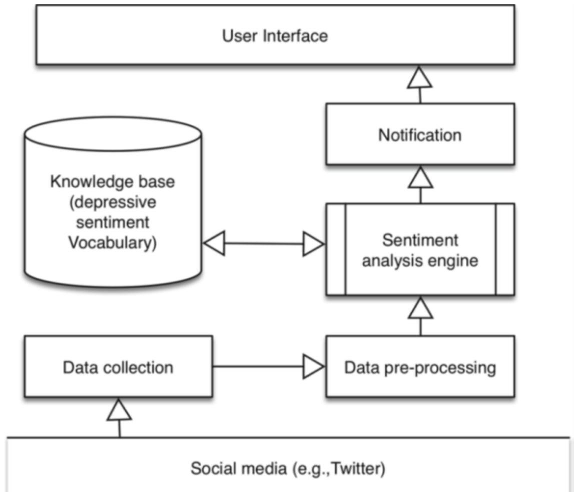

Fig. 1: Sentiment Analysis for Depression Detection on Social Networks [11]

3.2 Proposed Model

The system first collects data (tweets) from social media. The data was then pre-processed and analyzed by the system communicating with a knowledge base storing depressive sentiment vocabulary. The analysis result, for instance, positive cases regarding potential depression was then analyzed to determine the intensity of depression, afterwards assigning therapy recommendations or treatments, also, aware warnings to a designated person as a follow up (e.g., Human resources, social workers, the parents of children) through the depression monitoring portal.

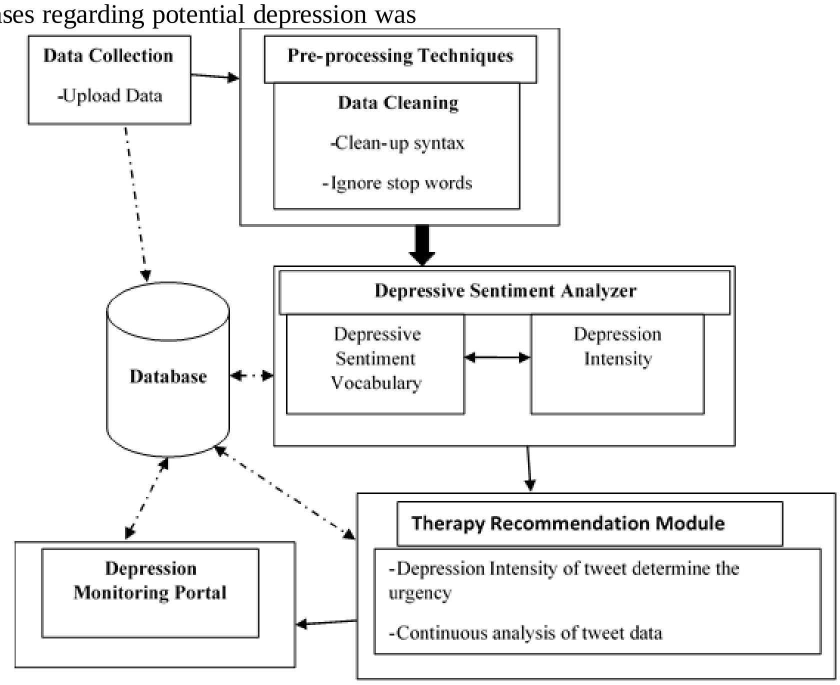

Figure 2: Proposed Depression Monitoring Model (DMM)

3.3 Data collection and Pre-processing

In this phase data was collected from twitter Application Programming Interface (API) and Depressive Sentiment Vocabulary (DSV) were gotten from *Myvocabulary.com* [15] and *Thesaurus.com* [16] The collected data served as datasets used to train and test the effectiveness of the DMM. My Structured Query Language (MySQL) server housed the knowledge base where DSV and the classification rules were stored. PHP served as the analytical tool. Valence Aware Dictionary for Sentiment Reasoning (VADER) model was used to determine the sentiment intensity. The sentiment intensity is to help analyze the magnitude of depression. Pre-processing of the data was done to improve the predictive accuracy of the dataset since it is susceptible to noise, missing and inconsistent data due to the nature of the data and human error.

For the purpose of this study five classification algorithms in machine learning such as Support Vector Machine, Bayes (Naive Bayes), Logistic Regression, k-Nearest Neighbor and decision tree were implemented on RapidMiner 9.3 for easy and fair comparison of each algorithm. In evaluating the performance of the classification algorithms, the model was built in RapidMiner 9.3 using the hold-out (66% training data) and 10-fold cross-validation evaluation methods on Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), k-Nearest Neighbor (k-NN), and Decision Tree (DT) algorithms to train and test the classifiers. After the training process, the values of correctly classified instances, time to learn, kappa statistics, specificity and sensitivity were computed to compare their performances in which two algorithms with leading performances were chosen.

Three phases were involved in building the model. These phases include pre-processing, processing and post-processing. The pre-processing phase involves feature transformation and extraction, the processing phase involves model generation and tuning based on the chosen algorithm; and post-processing phase involves knowledge representation. A schematic presentation of the depression monitoring model is shown in Figure 2.

4. Expected Outcome

Sentiment analysis technique on classification algorithms in machine learning to determine optimal algorithm among the five selected algorithms.

By utilizing machine learning techniques, the depression monitoring model will be implemented as a web-based Depression Monitoring System (DMS) that can identify depression risk group using tweets, as increase in words in tweets from the depressive sentiment vocabulary would lead to depression.

The findings will provide better insight into the process of adopting sentiment analysis technique in machine learning algorithms to solve depression problems, thus revealing how machine learning algorithm can be applied to real world problem.

5. Conclusion and Recommendations

Sentiment Analysis can help to monitor people’s mood. People with symptoms of depression have similar behavior that can be expressed in the phrases posted on social media. Thus, this useful information helps to determine users who have potential psychological disturbs such as depression. Also, authorized persons, relatives and medical personnel can have access to this information.In conclusion, this research proffers a web-based DMS that provides a solution to the problem of identifying depression, analyzing and recommending therapy to such individuals. With these solutions the aftermath of depression which is death will drastically reduce.

References

[1] H. Ritchie, and M. Roser, "Mental Health," 2018. [Online]. Available: https://ourworldindata.org/mental-health. .

[2] WHO, "Depression," 2018. [Online]. Available: World Health Organization: https://www.who.int/news-room/fact-sheets/detail/depression..

[3] WHO, "Mental Health: Suicide Data," 2018. [Online]. Available: World Health Organization: https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/.

[4] R. Feintzeig, "With Workplace Suicides Rising, Companies Plan for the Unthinkable," 2018. [Online]. Available: https://www.wsj.com/articles/with-workplace-suicides-rising-companies-plan-for-the-unthinkable-1516205932.

[5] S. Advanced, "expertsystem," 2017. [Online]. Available: https://expertsystem.com/social-media-data-mining/.

[6] J.R, Brubaker, F. Kivran-Swaine, L. Taber, and G.R. Hayes, "Grief-stricken in a crowd: the language of bereavement and distress in social media," in Proceedings of ICWSM, 2012, pp. 42-49.

[7] 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 Li, J., Cao, L., Wang, C., Tan, K.C., Liu, B., Pei, J., Tseng, V.S. (eds.) PAKDD LNCS, 2013, Vol. 7867, pp. 201-213.

[8] M. De Choudhury, S. Counts, and, E. Horvitz, "Social Media as a measurement tool of depression in populations," in Proceedings of the 5th Annual ACM Web Science Conference, 2013, pp. 47-56.

[9] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from twitter activity," in CHI'15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 3187-3196.

[10] M. Nadeem, "Identifying depression on Twitter," arXiv preprint:1607.07384, 2016.

[11] X. Tao, X. Zhou, J. Zhang, and J. Yong, "Sentiment Analysis for Depression Detection," Springer International Publishing, pp. 807-810, 2016.

[12] D.E. Losada, and F. Crestani, "A test collection for research on depression and language use," N. Q. Fuhr, Ed.) CLEF LNCS, Springer, Cham, vol. 9822, pp. 28-39, 2016.

[13] R. L. Rosa, D. Z. Rodriquez., G. M. Schwartz, I. de Campos Ribeiro, and G. Bressan, "Monitoring system for potential users with depression using sentiment analysis," in IEEE International Conference on Consumer Electronics (ICCE), 2016, pp. 381-382.

[14] V. Leiva, and A. Freire, "“Towards Suicide Prevention: Early Detection of Depression on Social Media," Springer International Publishing, pp. pp. 428-436, 2017.

[15] myvocabulary.com, "Depression Vocabulary Word List," 2016. [Online]. Available: https://myvocabulary.com/word-list/depression-vocabulary/.

[16] Dictionary.com, "Depression," 2016. [Online]. Available: http://www.thesaurus.com/browse/depression.