

Review: A survey on Semantic Sentiment Analysis using Machine Learning for Suicidal Tendency Prediction from Social Network

¹Ashrita Shivananda Hegde, ²Dhanushree B, ³Kavana L, ⁴Thanushree K A, ⁵Vimuktha E Salis,

⁵Associate Professor of Information Science and Engineering,

^{1,2,3,4}Student, Department of Information Science and Engineering, Global Academy of Technology Bangalore

Abstract :

With the increase in social networking spots, people are now more engaged in their virtual lives than ever ahead and at the same time, the number of information people put online is enormous, offering experimenters an inestimable tool for conducting exploration. People tend to put their studies online to partake with the whole world which also includes suicidal studies. Self-murder is a social problem and is a major concern in recent times. In this exploration paper, we substantially concentrated on Twitter which is one of the well-known networking spots. We espoused an approach of machine literacy and neural network for this exploration. Support Vector Machine (SVM) is one of the stylish machine literacy algorithms for textbook analysis and neural network is also well known for their performance incomplete cases. In the case of the neural network, we used three types of weight optimizers videlicet Limited- memory BFGS, Stochastic grade descent, and an extension of stochastic grade descent which is Adam to attain maximum delicacy.

Date of Submission: 15-04-2022

Date of acceptance: 30-04-2022

I. INTRODUCTION

Moment, people witness severe physical diseases and cerebral stress due to a variety of internal and external factors. Although depression is substantially planted in people in their 30s and 40s, it's frequently detected in kids due to academic stress and interpersonal relationship and in senior persons. In India, there are colorful social causes as well that promote self-murder attempts, the most common being dowry controversies. A lot of vituperative content, importunity, gibing, and cyberbullying affiliated conditioning has come veritably common on similar platforms which have dangerous goods on a person's internal and cerebral health. This can occasionally lead to mischievous and life-long traumatic goods on an existent. Our exploration pretends to collect the intimately available media dispatches of healthy and tone-diagnosed individualities which contain mixed Feelings so estimate the uprooted Twitter data and apply machine literacy classifiers similar to Naive Bayes, SVM, and deep literacy classifiers similar as LSTM-RNN to prognosticate depressive and anxiety tweets. Our results indicate that connections between appreciatively and negatively valanced terms give rise to a degree of balance that's more significantly advanced than in a null model where the affective structure is randomized and in a verbal birth, model capturing mind- the wandering in absence of suicidal creativity.

II. LITERATURE SURVEY

Ranjitha Korrapati [1] proposed creating chatbots based on artificial intelligence, but only 75% of them are accurate. The methodology used in this chatbots is machine learning and data analysis.

Data Analysis- use data mining ways to uncover retired trends in the dataset and Semantic Sentiment Analysis using Machine Learning for Suicidal Tendency Vaticination from social Networks to find the correlation between the variables, plot different graphs state-wise to uncover trends in self-murder rates and find the different factors behind self-murders. Strings (plain textbooks) cannot be reused by machines or deep literacy algorithms. Their raw input cannot be reused by these algorithms. This problem can be answered with the word embedding fashion which converts the string textbook into a numerical or vector format that the model can use. The perpetration involves a colorful corridor related to data reclamation, similar to getting and loading datasets into pandas' data frames for processing.

S.V. Georgakopoulos [2] introduced the implementation of CNN features involving FASTTEXT as a method of increasing classification accuracy through faster detection. The goal of fasttext is to efficiently learn word representations and categorize sentences.

J T Wolohan [3] proposed the use of word embedding for identifying depressive tweets from Twitter using classification tasks for depressive tweets. Word embeddings are a method for extracting features from the text that we can then use in a machine learning model to process text data. Syntactical and semantic information are preserved. These methods, which include Bag of Words (BOW), CountVectorizer, and TFIDF rely on the number of words in a sentence but do not save any syntactical or semantic information.

Bag of Words (Arc) Analysis- Several words associated with the subject matter are named. To identify the extent to which the textbook matches the words and the area as a whole, the named words and the textbook are vindicated against each other. The methodology used is NLP, Machine Learning, Deep Learning, and Naïve Bayes.

Kharde A.V [4] proposed the Textblob module for Sentiment Analysis, which allows for analyzing and extracting data from messages. Using it, you will be able to explore common natural language processing (NLP) tasks, such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and translation.

T. Zhang [5] proposed machine learning and deep learning approaches to different datasets using Recurrent Neural Network. Recurrent neural networks (RNN) are artificial neural networks that use sequential data or time-series data. They are widely used for ordinal or temporal problems, such as language translation, natural language processing (NLP), speech recognition, and image captioning; they appear in popular applications such as Google Translate, and voice search. They are recognized by their "memory" since they use information from previous inputs to influence current inputs and outputs. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of recurrent neural networks depends on the prior elements within the sequence.

Mc Adams D.P [6] built a cognitive network of conceptual associations for the suicidal prediction that uses cognitive processes to make decisions based on decisions and learn from those decisions. To achieve this, he has used the co-occurrence linguistic network and subject-verb-object linguistic network.

Mia Johnson Vioules [7] designed and evaluated a novel approach to monitoring the mental health of a user on Twitter using Natural Language Processing (NLP). By analyzing, understanding, and interpreting text and speech in the same way as humans, computers can understand.

James A McCart [8] proposed RNNs with an attention mechanism for relational encoding and used a Neural network to classify the sentiment expressed in suicide notes. RNNs work by storing the output of one layer and feeding it back into another layer's input in order to predict its output.

J. Jashinsky [9] used Machine Learning, Natural Language Processing, and Data Mining. They were designed to read the text and judge psychological weights using neural networks.

Rifat Zahan [10] developed a machine learning model to analyze suicidal risk using logistic regression. Logistic regression is a classification model used to give observations to a discrete set of classes such as suicide attempters or non-attempters. Logistic regression converts its

output using the more complex cost function known as the logistic sigmoid function to return a probability value. Y.W.NM Razali [11] designed a highly predictive model that supports clinical decisions related to the risk of suicidal behavior. Data Mining, Machine Learning, and Linear Regression methodologies were used.

Kharde A.V [12] presented a Machine Learning approach to predict suicide using Twitter data. Deep learning can interpret the risk of suicide by text classification.

Text Classification/Text Categorization is one of the important and typical tasks in ML. Using this technique, machines can identify organized groups of text by understanding its content.

K Venkateswara Rao [13] developed a variety of machine-learning algorithms and analyzed the data using WordNet semantics. A bag-of-words model is used for extracting features from the text used in machine learning algorithms.

D. Delgado-Gomez [14] proposed a model for predicting suicidal behavior using Twitter and Reddit datasets. Krauss MJ [15] used the TF-IDF algorithm to examine the feelings of the client for analyzing depression, self-harm, and suicidal ideation content on Tumblr. For information retrieval, TF-IDF is used. The TF*IDF algorithm is used to weigh keywords and assign importance to each keyword based on the number of times it appears in the document. Furthermore, it evaluates how relevant a keyword is across the web, referred to as corpus.

III. COMPARISON

| Authors | Methodology Used | Contribution |
|--------------------------|--|--|
| Ranjitha Korrapati [1] | Machine learning, And Data Analysis | AI-based chat bots have been developed and the accuracy is closed to only 75% . |
| S.V.Georgakopoulos [2] | Convolution Neural network(CNN), Fastext | Implementation of CNN features using the FASTEXT method and to increase the detection rate for accurate classification. |
| J T Wolohan [3] | NLP, Machine Learning, Deep Learning, Naïve Bayes | Presented a novel approach to word embedding for classification tasks to detect the depressive tweets from Twitter. |
| Kharde, A.V [4] | Textblob | Textblob module for Sentiment Analysis is used for extracting and analyzing data from messages |
| T. Zhang[5] | Recurrent Neural Network | Demonstrated machine learning and deep learning approaches on different datasets. |
| McAdams, D. P[6] | Co-occurrence linguistic network, Subject-verb-object linguistic network | Built and cognitive networks of conceptual associations |
| Mia Johnson Vioulès [7] | NLP, Machine learning text | Designed and evaluated a novel approach to monitoring the mental health of a user on Twitter using NLP |
| James A McCart [8] | Neural network, RCNN | Proposes RNs with an attention mechanism for relational encoding |
| J. Jashinsky [9] | Machine Learning Natural Language Processing Data Mining | The neural networks constructed were designed to read the text and infer psychological weights. |
| Rifat Zahan [10] | ML, Logistic, Regression, Decision tree | Developed a machine learning model to analyze suicidal risk. |
| Y. W. NM Razali[11] | Data Mining, Machine Learning, Linear Regression | Developed a highly predictive model that supports clinical decisions related to the risk of suicidal behavior. |
| Kharde,A.V [12] | Deep Learning, Machine | Presented Machine Learning approach to predict suicide. |
| K Venkateshwara Rao [13] | Machine Learning, vocabulary dictionary formation, Data extraction | Implemented multiple types of algorithms related to machine learning and by using WordNet semantically analyzed the data. |
| D. Delgado-Gomez [14] | Deep Learning, Machine learning | Predicting suicidal behavior using Twitter and Reddit dataset. |
| Krauss MJ, [15] | Classifier, TF-IDF Algorithm | Examines the feelings of the client as well as figures different outcomes like the client with greatest companions, top tweets, and so on. |

IV. APPLICATIONS

The detection of suicidal ideation has been enhanced through the application of several machine learning techniques. Data sources can also be adjusted to view relevant text research. Several domains are covered by the application, including questionnaires, electronic health records, suicide notes, and online user content. Through our use of networks, we can decipher the meaning of online posts, provide early warnings, and even alleviate a person's suicidal intentions.

V. CONCLUSION

In this article, we have reviewed the research done in the area of suicide prediction. Even though many successful models have been made with high accuracy, further improvements are possible. Several hidden results were revealed in the data analysis, such as the connection between introversion and depression and that male teens are more likely to commit suicide. It will also assist government agencies working towards suicide prevention and counseling to take effective steps forward and provide insight into improvement areas. It has been demonstrated that machine learning algorithms, such as gradient boosted decision trees and neural networks, consistently outperformed other algorithms and had the highest accuracy and precision. This article reports the development of a model that can predict suicide attempts or suggest suicide causes from readily available social media data. This model can be used to effectively predict suicidal behavior in a sensitive, specific, and timely manner.

REFERENCES

- [1]. Ranjitha Korrapati, Kranthi Nuthalapati, S. Thenmalar, "A Survey Paper on Suicide Analysis", *International Journal of Pure and Applied Mathematics*, Volume 118 No. 22 2018, 239-244.
- [2]. S. V. Georgakopoulos, A. G. Vrahatis, S. K. Tasoulis, V. P. Plagianakos. "Convolutional Neural Networks for Toxic Comment Classification." *arXiv:1802.09957v1 [cs.CL]*, 27 Feb 2018.
- [3]. J.T. Wolohan, Misato Hiraga, "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP", Department of Information and Library Science Indiana University - Bloomington.
- [4]. Kharde, A.V., & Sonawane, S.S. —Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, Volume 139, No.11, 2016.
- [5]. T. Zhang, <https://doi.org/10.1016/j.invent.2021.100422>
- [6]. Received 16 February 2021; Received in revised form 2 June 2021; Accepted 22 June 2021. [6]. McAdams, D. P. The psychology of life stories. *Rev. Gen. Psychol.* 5, 100–122 (2001).
- [7]. Mia Johnson Vioulès, Bilel Moulahi, Jérôme Azé, Sandra Bringay. Detection of Suicide Related Posts in Twitter Data Streams. *Ibm Journal of Research and Development*, IBM Corporation, 2018, 62 (1), pp. 7:1–7:12. [ff10.1147/JRD.2017.2768678ff](https://doi.org/10.1147/JRD.2017.2768678ff). [fflirmm-01633317ff](https://doi.org/10.1147/JRD.2017.2768678ff)
- [8]. James A McCart, Dizon K Finch, Jay Jarman, Edward Hickling, Jason D Lind, Matthew R Richard, Donald J Berndt, and Stephen L Luther. Using ensemble models to classify the sentiment expressed in notes. *Biomedical informatics insights*, 5:BII–S8931, 2012.
- [9]. J. Jashinsky, S. H. Burton, C. L. Hanson et al., "Tracking suicide risk factors through Twitter in the US," *Crisis*, vol. 35, no. 1, pp. 51–59, 2014.
- [10]. Rifat Zahan; Ian McQuillan; Nathaniel Osgood, "DNA Methylation Data to Predict Suicidal and Non-Suicidal Deaths: A Machine Learning Approach", 2018 IEEE International Conference on Healthcare
- [11]. Y. M. Tai and H. W. Chiu, "Artificial neural network analysis on suicide and self-harm history of Taiwanese soldiers," in *Second International Conference on Innovative Computing, Information, and Control (ICICIC 2007)*, pp. 363–363, Kuma-moto, Japan, 2007, IEEE.
- [12]. Kharde, A.V., & Sonawane, S.S. —Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, Volume 139, No.11, 2016.
- [13]. K Venkateswara Rao, B. O'Dea, S. Wan, P. J. Batterham, A. L. Callear, C. Paris, and H. Christensen. Detecting suicidality on Twitter. *Internet Interventions*, vol. 2, pp. 183–188, 2015.
- [14]. D. Delgado-Gomez, H. Blasco-Fontecilla, F. Sukno, M. Socorro Ramos-Plasencia, and E. Baca-Garcia, "Suicide attempters classification: toward predictive models of suicidal behavior," *Neurocomputing*, vol. 92, pp. 3–8, 2012.
- [15]. Krauss MJ, Cavazos-Rehg PA, Connolly S, Sowles SJ, Bharadwaj M, Rosas C, et al. An Analysis of Depression, Self-Harm, and Suicidal Ideation Content on Tumblr. *Crisis* 2016 Jul 22:1-9