

# Suicide Text Detection using Machine Learning

Tanmay Khopkar<sup>1</sup>, Akshay Khane<sup>2</sup>, Prathamesh Dhumal<sup>3</sup>, Mrs. Snehal R. Shinde<sup>4</sup>

<sup>1,2,3</sup>Department of Computer Engineering, Pillai HOC College of Engineering and Technology, Rasayani, Maharashtra, India.

<sup>4</sup>Assistant Professor Snehal R. Shinde, Department of Computer Engineering, Pillai HOC College of Engineering and Technology, Rasayani, Maharashtra, India

\*\*\*

**Abstract** - Social media network like twitter is a communication channel which enables it's users to broadcast their thoughts through brief text updates. Different people have different thoughts which they often post on Twitter. Close to 800,000 people commit suicide every year, which is 1 person every 40 seconds. Vulnerable accounts, usually post about their depressive situations on social media via twitter prior to act. Raw but useful data is available on social media, these account's social media activity can be observed by mining the data using different Machine Learning techniques. Convolutional Neural Network has been used to classify the text in these raw data. This result, to the best of our knowledge, has the best performance for classifying data from social media.

**Key Words:** Text Classification, Machine Learning, CNN, Social Media, Suicide.

## 1. INTRODUCTION

Social media is a type of online intercommunication platform that allow its users to share different kinds of media publicly. The media can be in the form of photos, videos, text messages, etc. There are different types of social media platforms that are very popular.

Twitter is one of the social media platforms that is used by millions of people. Initially, Twitter only allowed sharing of texts called as "Tweets", but, it now also allows photos and videos to be shared. Twitter can be used in many ways by its users for e.g. advertising products, sharing thoughts, sharing information about a campaign, an event etc. Twitter is generally used by its users for sharing their thoughts. People usually share their personal, political, professional thoughts onto the platform.

Due to the rise of internet usage, number of people using the internet has drastically increased over time. As a result, the number of people using social media, as in our case, Twitter, has also increased and so is the content on social media. People share their opinions, feelings, thoughts etc. on Twitter.

Feeling or emotions shared by the users can be depressive or joyful. The publisher of the depressive text can either be just sharing his/hers feeling or it can be a cry for help. In an extreme case, the user can be suicidal i.e. likely to commit

suicide. Our aim is to detect such extreme cases of suicidal individuals using their tweets they post onto Twitter.

## 2. RELATED WORK

Large amount of studies have been done on recognising the suicidal tendency of an individual in many fields like medical fields using the resting-state of the heart rate, psychology using the thoughts and feeling of the individual.

Machine learning fields also have been used to research and determine the suicidal tendency of an individual using the content that individual posts on the social media. Text Classification method of machine learning is used to determine the suicidal tendency. Besides N-gram features, knowledge based features, syntactic features, models for cyber suicide detection also used regression analysis ANN and CRF. Mulholland and Quinn extracted the vocabulary and verbal features to determine the suicidal or non-suicidal sentences. Huang et al. built a psychological lexicon dictionary and used a SVM classifier. [11]Chatopadhyay proposed a mathematical feed-forward multilayer neural network. [9]Delgado-Gomez et al and [9]Pestian et al. compared the performance of different multivariate complexity techniques.

## 3. METHODOLOGY

### 3.1 Pre-Processing

Data pre-processing is a process that converts the raw, real-life data in an understandable format. The data in the raw form is often incomplete and isn't understandable. So, in order to process the data and get the results that we desire, we need to perform data pre-processing on the raw data set so that actual results can be obtained.

The dataset that we have contains 5 columns, namely, id's, date, flag, user and tweet. But, among these columns we are only concerned with the tweet column, so, we drop all the columns except tweets column. The tweets are not present in only text format. The tweets contain special characters, hyperlinks, numbers and punctuations. We need to remove all of these and converts the contents of the column in a normalised form.

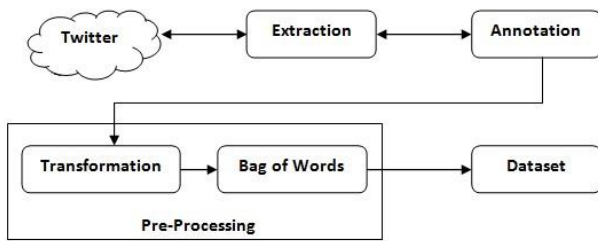


Fig -1: Data Pre-processing

After removing all the special characters, hyperlinks, numbers and punctuations, we now remove the connectors or stopwords for the tweets. The next step is tokenization, i.e. after the connectors have been removed, we need to split the string that is tweet in our case into a list tokens. From the tokens that we have now, there may be multiple tokens that are different forms of same word, so we use Porters stemmers from nltk library. After this, we separate the tokens into positive and negative words and visualize them using the Wordcloud. From the dataset, we visualize the words in 3 different groups, namely

- a) Words that are most commonly used in the tweets,
- b) Positive words that are most commonly used in the tweets, and
- c) Negative words that are most commonly used in the tweets.

### 3.2 Classification

Classification is the process of clubbing different objects with same properties into classes, in our case, based on the positive and negative words, determining the tweets is suicidal or non-suicidal. Among all of the classifiers present, we use Convolutional Neural Network (CNN) for classifying the tweets into suicidal or non-suicidal.

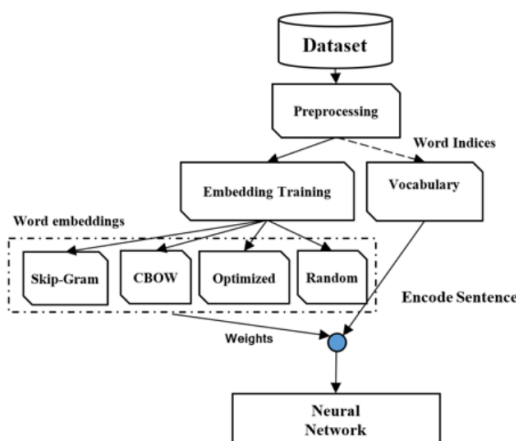


Fig -2: Convolutional Neural Network (CNN) Classifier

Neural word embedding have become popular for modelling the individual words and the interactions between them. It is used to serve various purposes like text classification. It uses distributed word representation where a single word isn't represented by a single neurons but a word is represented by numerous neurons. This helps the distributed representation system learn about the general concepts of language and not just the independent word representations. The most commonly used word embedding system was developed by Google called as Word2Vec which uses two neural network architectures namely,

- a) Continuous Bag of Words (CBoW) and,
- b) Skip-Gram (SG)

The Google Word2Vec basically converts the words into vectors using some pre-defined formula. It can even perform vectorization even for the unknown words. After the vectorization of the text is done, the two main processes for classifying the sentence, the tweet in our case are,

- a) Convolutional Filter or Kernal- The process in which a filter is applied on the vectorised values of the words using a sliding window of a pre-defined size and the new vectors are generated.
- b) Pooling- The process of sampling the new vectors created by applying the convolutional filter and combining them by taking the maximum or the average value.

Once the vectorization of the words and then sampling of the vectors is done, the vectors values are then given a inputs to the neurons of the Convolutional Neural Network (CNN). The are appropriately weighted and given to the neurons for processing. Rectified Linear Unit (ReLU) is given as activation functions to the neurons. Based on the inputs, the weights assigned to the inputs and the activation function assigned to the neurons, the network segregates the inputs into classes.

### 4. CONCLUSIONS

Our systems elaborated based on real time Twitter data. Among many widely used text classification models, based on this study CNN suicidal detection model shows better performance. This model shows better optimization on iterative solutions. This system is primarily focuses on suicidal people by mining the database system so that tweets related to suicide will get detected.

### FUTURE SCOPE

When our system detects the real time suicidal text we can make the program evaluate in such a way that we can find the geographical location of that particular tweet. There are many NGO's and anti suicidal organizations work against suicide.

we can inform these NGO's about such accounts so as to take further appropriate action. This system is not only can be implemented on Twitter but also other social media platforms like Facebook, Instagram, and WhatsApp.

## REFERENCES

- [1] Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015a. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses. In *Computational Linguistics and Clinical Psychology*, pages 1–10M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [2] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (5th ed.)*, 5 edition. American Psychiatric Publishing, Washington.K. Elissa, "Title of paper if known," unpublished.
- [3] *Depression and Other Common Mental Disorders: Global Health Estimates*. World Health Organization, 2017.
- [4] [BasicstructureaboutCNNWikipedia//en.wikipedia.org/wiki/Convolutionalneuralnetwork](https://en.wikipedia.org/wiki/Convolutional_neural_network).
- [5] E. Okhapkina, V. Okhapkin, and O. Kazarin, "Adaptation of information retrieval methods for identifying of destructive informational influence in social networks," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 87–92, Taipei, Taiwan, 2017, IEEE.
- [6] M. Mulholland and J. Quinn, "Suicidal tendencies: the automatic classification of suicidal and non-suicidal lyricists using NLP," in *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 680–684, Nagoya, Japan, 2013.
- [7] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, and T.Zhu, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, pp. 844–849, Bali, Indonesia, 2014, IEEE
- [8] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, and A. Leenaars, "Suicide note classification using natural language processing: a content analysis," *Biomedical Informatics Insights*, vol. 3, pp. 19–28, 2010.
- [9] 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.
- [10] Research Article Based on "Supervised Learning for Suicidal Ideation Detection in Online User Content" by Shaoxiong Ji, Celina Ping Yu, Sai-fu Fung, Shirui Pan, and Guodong Long.
- [11] S. Chattopadhyay, "A study on suicidal risk analysis," in *2007 9th International Conference on e-Health Networking, Application and Services*, pp. 74–78, Taipei, Taiwan, 2007, IEEE.