International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-08, Issue-01, APR 2022

Content Analysis of Messages in Social Networks, Identification of Suicidal Types

[1]Prof.Satish Manje, [2]Mr.Shubham Dhawale, [3]Mr.Acchutam Kulthe, [4]Mr.Rohit Ghoshtekar
[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. JondhleCollege of Engineering & Technology, Asangaon, Maharashatra, India. *[1]satishmanje93@gmail.com, [2]shubhamdhawale05@gmail.com,[3]kacchutam@gmail.com, [4]rohitghoshtekar@gmail.com*

Abstract- This project describes a content analysis of text to spot suicidal tendencies and kinds. This article describes the way to make a sentence classifier that uses a neural network created using various libraries created for machine learning within the Python programing language. Attention is paid to the matter of teenage suicide and «groups of death» in social networks, the look for ways to prevent the propaganda of suicide among minors. Analysis of existing information about so-called "groups of death" and its distribution on the Internet.

Keywords- LSTM- CNN Classifier

I. INTRODUCTION

Suicide ideation is stated as a desire to put life on hold, which can range from sadnessto planning a suicide attempt to an obsessive obsession with self-destruction. Suicide planners and suicide completers are two types of people who are at danger. In research, the link between the two groups is frequently discussed communities. The majority of persons who experience suicidal thoughts, according to certain studies, do not really commit themselves. For example, much of the commonly mentioned depression, hopelessness, and frustration associated with suicide are thought to be predictors of suicide thoughts rather than the passagefrom ideation to attempt. Self-immolation is defined a desire to put life on hold that can range from sadness to planning a suicide attempt to an obsessive obsession with self-immolation. Suicide planners and suicide completers are two types of people who are at danger. The link between the two groups is a major matter of debate in most research communities. According to certain surveys, the vast number of individuals who consider self-immolation do not go through with it. Pompili et al., in opposition, it show that he/she is both a suicide planner and a suicide attempter might to be very similar to "many variables thought to be suicidal risk factors conduct." Early detection of suicidal ideation has been created and implemented as a national strategy for preventing suicide in WHO nations in order to work towards a worldwide market with the average goal of reducing suicide rates will have decreased by 10%.. The increased word embeddings success are evident in risk assessments.

N-gram analysis: We use n-gram analysis, so can show that suicidal thoughts and diminished social interaction are frequently mentioned in suicide-related forums. The move to social ideation is connected to a variety of psychological stages, including increased self-centered attention is a sign of hopelessness. dissatisfaction, anxiety, or loneliness.

II. AIMS AND OBJECTIVE

a) Aim

The goal of the research is to communicate data-driven information's of suicide ideation in social media forums utilizing effectual deep learning architecture.

b) Objective

On a real-world dataset, compare it to CNN in terms of power and potential & LSTM deep learning algorithms, as well as four classic machine learning classifiers (NB, RF, SVM, & XG-Boost).

III. LITERATURE SURVEY

Paper 1: Addressing Suicidal Thoughts and Behaviors in Substance Abuse Treatment: Information You Need ToKnow:

Suicidal ideation and behavior is also a strong indicator of other coexisting disorders that must be investigated, identified, and addressed in order to improve substance misuse treatment outcomes. As a result, substance misuse treatment providers should be ready to frequently collect information from, refer, or assist in the cure of clients who are at risk of suicide. Suicidal cognitions are also an indication of other co-occurring illnesses that must beinvestigated, diagnosed, and treated in order to improve substance misuse treatment outcomes.

Paper 2: Scikit-learn: Machine Learning in Python:

It has few prerequisites and is released underneath the BSD License, making it appropriate for use in both commercial

 DOI : 10.35291/2454-9150.2022.0173

andacademic settings.

Paper 3: National suicide prevention strategies: progress, examples and indicators:

Scikit-learn is Python package that combines a variety of cutting-edge machine learning approaches targeting in supervised and unsupervised medium-scale problems. This package seeks to make machine learning accessible to non-experts by using a comparatively elevated language. The following features are prioritized: simplicity of use, efficiency, documentation, and API integrity are the aspects that are prioritized. It has few prerequisites and is distributed underneath the BSD License, making itsuitable for commercial and academic use.

Paper 4: Differentiating suicideattempters from suicide ideators: a critical frontier for suicidology research:

This paper proposes first multi-indicator method for determining the location where a tweet was created in addition location of the user's residence.

IV. EXISTING SYSTEM

The current system examines theconnection between social media sites and and suicide-ideation using a data set acquired from a controlling social networking service (SNS). This approach addresses limitations very less. First, anwhole social network of users is available, where a link between two users represents explicit bidirectional friendship initiated by both users. Some users have quite a many numbers of friends, as in general social networks. Second, for the same reason, we can exactly calculate the number of triangles for each user.

DISADVANTAGES OF EXISTING

SYSTEM: A function of mixing relevant to this study is user-defined community. It is high reliance on a proper presentation ofyour data.

This means that legit model is not useful tool unless you have already identified all the important independent variables

V. COMPARATIVE STUDY

SR NO.	PAPER TITLE	UTHORNAME	ATION	PURPOSE	ECHNOLOGY
1.	Addressing Suicidal Thoughts & Behaviors in Substance Abuse Treatment: InformationYou Need To Know	Bryabrina T.V., Gibert A.I., ShtrahovaA.V	SAMHSA, 2015	Suicidal thoughts are also astrong indicator ofother coexistingproblems that mustbe investigated,recognised, and treated in order toimprove substancemisuse treatment outcomes.	TIP (Treatment Improvement Protocols)
2.	Scikit-learn: Machine Learning in Python	Pedregosa F., Varoquaux G., GramfortA., Michel V	Reasearch gate, 2012	It has few dependencies &is provided under the BSD licence, making it suitable for usage in both academic & commercialenvironments.	Scikit-learn
3.	National suicide prevention strategies: progress, examples &indicators	World Health Organization	WHO,2019	To accelerate the implementation of suicide prevention, a national strategy and plan of action are being prepared.	'LIVELIFE' national suicide preventionstrategy
4.	Differentiating suicideattempters from suicide ideators: a critical frontier for suicidology research	David Klonsky ,Alexis MMay	Wiley Online Library,2013	This paper proposes firstmulti-indicator method for determining the location where atweet was created as well as location of the user's residence.	Ideation-to-action framework

Table 1: Comparative Study

VI. PROBLEM STATEMENT

Detecting suicide-ideation and mental illness in the social media content is technically an autonyms task of text tagging. Conduct pulverized suicide autonyms and classifications of multiple mental health issues, which are naturally regarded as multi-class classification. For pulverized suicide risk, the risk levels include none, low, moderate, and severe risk, while mental health classification, specific mental disorders are depression, anxiety, bipolar, and so on. And two subtask for specific settings of datas in social content, post-level and user-level classification.

VII. PROPOSED SYSTEM

The study Experience of content analysis of suicidal statements on the Internetweb of persons with different level of suicidal activity collects data from pages List individuals who have committed suicide or are potential

 DOI : 10.35291/2454-9150.2022.0173

suicides. By analysing the collected information, a program called Text Analyst explores the causes of suicidal behaviour and their feelings. The purpose of the current study is to classify sentences into suicidal-ideation and non-suicidal-ideation using a neural network. In our system, according to random text, it is necessary to determine whether it is suicidal or not, i.e. to find the solution ofits binary classification. Classification isthe distribution of data by parameters.

VIII. ALGORITHM

The general idea of working of current system algorithm is given as follow:

Step 1 : Start
Step 2 : Collect the dataset. **Step 3:** Preprocess the dataset. def preprocess_tweet(text): text → re.sub('<[^>]*>', '', text) emoticons→ re.findall('(?::|;|=)(?:-)?(?:\)|\(|D|P)', text) lowercase_text→ re.sub('[\W]+', ' ', text.lower()) text→ lowercase_text + ' '.join(emoticons).replace('-', '') return text tqdm.pandas() path → settings.MEDIA_ROOT + "\\" +"data.csv" df → pd.read_csv(path) df['tweet'] → df['tweet'].progress_apply(preprocess_tweet)
Step 4 : Perform Text Extraction andtokenization on Data. stop_words→ set(stopwords.words('english')) tokenized → sent_tokenize(tweet)for i in tokenized: wordsList → nltk.word_tokenize(i) wordsList → [w for w in wordsList if notw in stop_words] tagged → nltk.pos_tag(wordsList) print(tagged)
Step 5: Perform text categorization on data.Extract words that represent mentalhealth and store it in Mental Health file and store suicidal words in Suicide watch file.Analyze the results using wordCloud library.
Step 6: Split the input into two sections: training and testing using Model_Selectionin sklearn library.Using numpy arrays. label train as 0 and test as 1.
Step 7: Train the model using convolutional Neural Network(CNN).Using Sigmoid activation function, for the negative and positive thoughts in the review, output a number in middle of 0 and 1.
Algo CNN:
X_train→ sequence.pad_sequences(X_train, maxlen → max_words) X_test → sequence.pad_sequences(X_test,maxlen → max_words) model → Sequential() model.add(Embedding(top_words, 32, input_length → max_words)) model.add(Conv1D(32, 3, padding →'same', activation → 'relu')) model.add(MaxPooling1D()) model.add(Flatten()) model.add(Dense(250, activation→ 'relu'))model.add(Dense(1, activation →'sigmoid')) model.compile(loss → 'binary_crossentropy', optimizer='adam',metrics → ['accuracy']) model.summary() model.fit(X_train, y_train, validation_data → (X_test, y_test), epochs → 10, batch_size → 128, verbose → 2) scores → model.evaluate(X_test, y_test,verbose → 0) return (scores[1] * 100)
Step 8: Test and run the model for input twitter text and calculate the predicted results and probability of input text as suicidal or non-suicidal.
Step 9: Stop

IX. MATHEMATICAL MODEL

1. Long Short Term Memory

At a high-level LSTM works very much like an RNN cell. The LSTM network'sinternal workings are shown here. The LSTM is made of three sections, as illustrated in the diagram below, each of which serves a different function. The first section decides whether the information from the preceding timestamp should be retained or ignored. The cell tries to learn new knowledge from the second section's input. Finally, the cell passes updated information from current to next timestamp in the third component. The gates are the three components of an LSTM cell. The Forget gate is the firstpart, the Input gate is the second, and Output gate is the final part.

2. Multivariate Linear Regression

The simple regression linear modelrepresents a straight line meaning y is function of x. When we have extra dimension (z), the straight line become plane. Here, the plane is function that expresses y as a function x and z. The linear regression equation can now be expressed as:

$$y = m1.x + m2.z + c \quad y = \beta0 + \beta1.x1 + \beta2.x2$$

If there are three variables as inputs? Human visualizations can only three dimensions. In the machine learning world, There could be an limitless number of dimensions. The equation for a model with three input variables can be written down as:

$$y = \beta0 + \beta1.x1 + \beta2.x2 + \beta3.x3$$ Below is the generalized equation for themultivariate regression model-

$$y = \beta0 + \beta1.x1 + \beta2.x2 + \ldots.. + \beta n.xn$$ The number of parameters that areindependent is represented by n, the coefficients are represented by 0 n, and theindependent variable is represented byx1~xn.

3. Convolutional Neural Network

CNNs are similar to classic ANNs in that they are made of neurons that learn to optimize themselves. Each neuron still receive an input and conduct an action which is the foundation of innumerable artificial neural networks. The entire network will still express a singleperceptual scoring function from the input raw picture vectors to the final output of the class score (the weight). All of the normal ANN tips and methods will stillapplicable in the final layer, which will contain the lost functions associated with the classes.

4. Convolutional Layer

It is a component of the CNN neural network, which was originally created for image recognition with high performance. The LSTM model extracts each feature sequence after that

$$H = [h1, h2, h3, ..., hT]$$

T where ht is an m-dimensional vector ofthe hT word, T is the number of LSTM expansion steps equal to the

 DOI : 10.35291/2454-9150.2022.0173

length of the text sequence in a text sequence.

CNN input matrix H RmT has fixed-lengthinputs, thus each input length is standardized to T by clipping longer words and padding short sentences with zero. F Rjk is the convolutional filter, with j denoting the number of words in window and k being the dimension of word embedding vector. The convolutional filter F = [F0, F2,…, Fm1] will generate one value at time step t, as shown below. Equation:

Fl = ReLU[(∑ hl+i Fi) + b]

This single filter's parameters are F and b, with b being a bias. Finally, non-linearity is removed by applying the ReLU activation function to a feature map. Itsmathematical formula is as follows:

F(x) = max(0, x)

The pooling layer's goal is to lower the dimensionality of each rectified feature map while keeping the most important data. It lowers the number of values & calculations in the network, allowing for better control of over-fitting. In our research, we make use a maximum pooling approach to represent the most important information in each feature map. Layer should be flattened. The goal of the CNN flatten layer is to convert a column a vector derived from a collection of feature that can be fed into the classification task's neural network.

Flattening = pooled.reshape

X. SYSTEM ARCHITECTURE

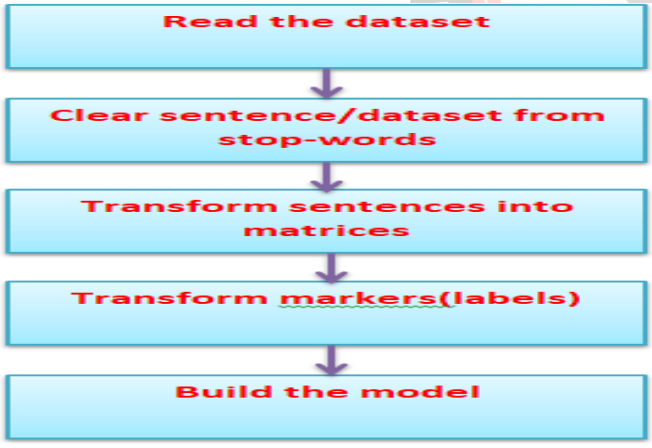

Fig. 1: System Architecture

XI. ADVANATGES

1) The collected dataset consists of two columns sentence and label, where sentence is a column with sentences, and label is column with the values suicidal/non-suicidal.

2) NLTK library to find a solution words do not carry any semantic load and symbols.

3Tokenizer library was transformsentences into arrays.

XII. DESIGN DETAILS

Fig. 2: Result of the entered tweet whether it is suicidal or not , along with its accuracy

XIII. CONCLUSION

Thus, We have attempted to implement the paper "A. A. Shvetsova and M. A. Antropova, "Content Analysis of Messages in Social Networks, Identification of Suicidal Types," 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering *(EIConRus")*, IEEE 2020 and The conclusion, according the execution, is for analyzing the information regarding suicidal ideation. This neural 9 network can be applied to scan content on social media sites and filter posts with suicidal undertones, preventing the spread and encouragement of suicide among young people who couldn't envisage life without social media.

Also, if a parent does not want to violate the privacy of their child's personal messages but is concerned about him, this project 10 can be used for parental control: all that is required is for the parent to collect gather information and submit it to the software. Also, this neural network can be put for use to block communities that contain a suspiciously large number of posts with suicidal content.

XIV. REFERENCE

[1] A. A. Shvetsova and M. A. Antropova,"Content Analysis of Messages in SocialNetworks, Identification of SuicidalTypes," 2020 IEEE Conference of RussianYoung Researchersin Electrical and Electronic Engineering *(EIConRus)*, 2020,pp.81-83,doi: 10.1109/EIConRus49466.2020.9038911.

[2] "Marks, M. Artificial Intelligence Based Suicide Prediction". Yale J. Health Policy Law Ethics 2019. Forthcoming. [Google Scholar]

[3] "Beck, A.T.";"Kovacs", "M.; Weissman", A. Hopelessness and suicidal behavior: An overview. JAMA 1975, 234, 1146–1149. [Google Scholar] [CrossRef][PubMed]

[4] "Silver","M.A.; Bohnert","M.; Beck, A.T.";"Marcus, D." Relation of depression of attempted

 DOI : 10.35291/2454-9150.2022.0173

suicide and seriousness of intent. Arch. Gen. Psychiatry 1971, 25, 573–576. [Google Scholar] [CrossRef] [PubMed]

[5]"Klonsky, E.D".;"May, A.M." Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. Suicide Life-Threat. Behav. 2014, 44, 1–5. [Google Scholar] [CrossRef] [PubMed]

[6] Pompili, M.; Innamorati, M.; Di Vittorio, C.; Sher, L.; Girardi, P.; Amore, M. Sociodemographic and clinical differences between suicide ideators and attempters: A study of mood disordered patients 50 years and older. Suicide Life-Threat. Behav. 2014, 44, 34–45. [Google Scholar] [CrossRef] [PubMed]

[7] DeJong, T.M.; Overholser, J.C.; Stockmeier, C.A. Apples to oranges?: A direct comparison between suicide attempters and suicide completers. J. Affect. Disord. 2010, 124, 90–97. [Google Scholar] [CrossRef] [PubMed]

[8] De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; Kumar, M. Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San José, CA, USA, 9–12 December 2016; ACM: New York, NY, USA, 2016; pp. 2098–2110. [Google Scholar]

 DOI : 10.35291/2454-9150.2022.0173