

Detection of Suicidal Ideation in Social Media Through Content Analysis

¹ K. Divya, ² Kona Namratha Blessy, ³ Yarramsetty Lalasa Navya, ⁴ Jonnala Kumar Swamy, ⁵ Sammeta Manikanta

¹ Assistant Professor, ^{2,3,4,5} UG Students

Department of CSE-AI & ML, Sri Vasavi Institute of Engineering & Technology, Nandamuru, Andhra Pradesh, India

ARTICLE INFO

Article History:

Accepted: 15 April 2024

Published: 27 April 2024

Publication Issue :

Volume 11, Issue 2

March-April-2024

Page Number :

862-869

ABSTRACT

This article describes content analysis of text with to identify suicidal tendencies and types. This article also describes how to make a sentence classifier that uses a neural network created using various libraries created for machine learning in the Python programming language. Attention is paid to the problem of teenage suicide and «groups of death» in social networks, the search for ways to stop the propaganda of suicide among minors. Analysis of existing information about so-called «groups of death» and its distribution on the Internet.

Keywords : Content Analysis; Social Networks; Neural Networks; Machine Learning; Groups of Death; Python

I. INTRODUCTION

In 2016, the media noted an increase in the number of teenage suicides, linking it to thematic groups on social networks. The study «Groups of death» - from play to moral panic» describes how it attracted children of 12-16 years in such groups. Society actively discussed the so-called game «Blue whale», suddenly became popular among teenagers. The study «Experience of content analysis of suicidal statements on the Internet of persons with different levels of suicidal activity» [2] collects data from the pages of people who have actually committed suicide or are potential suicides. By analyzing the collected information, program called TextAnalyst explores the

causes of suicidal behavior and their feelings. The aim of the current study is to classify sentences into suicidal and non-suicidal using a neural network.

This article describes content analysis of text with to identify suicidal tendencies and types. By analyzing the collected information, program called TextAnalyst explores the causes of suicidal behavior and their feelings. The aim of the current study is to classify sentences into suicidal and non-suicidal using a neural network.

To describes how to make a sentence classifier that uses a neural network created using various libraries created for machine learning in the Python programming language. In our case, according to random text, it is necessary to determine whether it is

suicidal or not, i.e. to solve the problem of its binary classification. Classification is the distribution of data by parameters.

II.RELATED WORK

Investigating Suicide-Related Subliminal Messages on Instagram

Text data analysis of social media is becoming more and more important since it includes the Suicide is the second-leading cause of death among 15-29-year-olds and Instagram is one of the most popular and fastest-growing social media platforms among this age group. A previous study presented preliminary evidence for suicide-related "subliminal messages" on Instagram, defined as very brief presentations of suicide-related content in video posts that users have no conscious awareness of. Aim: A systematic quantitative study was pending. Method: We conducted a quantitative content analysis of 100 Instagram video posts. A frame-by-frame coding procedure allowed for an assessment of whether suicide-related content was depicted in very brief segments, even when this content could not be consciously recognized when watched at regular speed. Results: Analysis indicates that a substantial amount of suicide-related content is presented in very brief shots. We identified 67 very brief shots that appeared in 21 video posts. Of interest, 13 of these video posts presented more than one very brief suicide-related shot. Limitation: The subjective threshold of conscious awareness differs inter-individually. This complicates the operationalization of subliminal messages. Conclusion: Subliminal messages are ethically highly problematic. There is a need for a greater awareness of possible suicide-related subliminal messages on Instagram.

Suicide on Instagram - Content Analysis of a German Suicide-Related Hashtag

Approximately half of all posts included words or visuals related to suicide. Cutting was by far the most prominent method. Although sadness was the

dominant emotion, self-hate and loneliness also appeared regularly. Importantly, inconsistency - a gap between one's inner mental state (e.g., sadness) and one's overtly expressed behavior (e.g., smiling) - was also a recurring theme. Conversely, help-seeking, death wishes, and professional awareness-intervention material were very rare. An explorative analysis revealed that some videos relied on very fast cutting techniques. We provide tentative evidence that users may be exposed to purposefully inserted suicide-related subliminal messages (i.e., exposure to content without the user's conscious awareness).

Content shared on social media for national cancer survivors day 2018

Studies estimate that the number of cancer survivors will double by 2050 due to improvements in diagnostic accuracy and treatment efficacy. Despite the growing population of cancer survivors, there is a paucity of research regarding how these individuals experience the transition from active treatment to long-term surveillance. While research has explored this transition from more organized venues, such as support groups for cancer survivors, this paper explores the discourses surrounding cancer survivorship on social media, paying particular attention to how individuals who identify as cancer survivors represent their experience.

Ebola virus disease and social media: A systematic review

Twelve articles were included in the main analysis: 7 from Twitter with 1 also including Weibo, 1 from Face book, 3 from YouTube, and 1 from Instagram and Flickr. All the studies were cross-sectional. Eleven of the 12 articles studied ≥ 1 of these 3 elements of social media and their relationships: themes or topics of social media contents, meta-data of social media posts (such as frequency of original posts and reposts, and impressions) and characteristics of the social media accounts that made these posts (such as whether they are individuals or institutions). One article studied how news videos influenced Twitter traffic. Twitter content analysis methods

included text mining ($n = 3$) and manual coding ($n = 1$). Two studies involved mathematical modeling. All 3 YouTube studies and the Instagram/Flickr study used manual coding of videos and images, respectively.

Twitter as a Tool for Health Research: A Systematic Review

We identified a new taxonomy to describe Twitter use in health research with 6 categories. Many data elements discernible from a user's Twitter profile, especially demographics, have been underreported in the literature and can provide new opportunities to characterize the users whose data are analyzed in these studies. Twitter-based health research is a growing field funded by a diversity of organizations. Public health implications. Future work should develop standardized reporting guidelines for health researchers who use Twitter and policies that address privacy and ethical concerns in social media research.

III.PROPOSED SYSTEM

This article describes content analysis of text with to identify suicidal tendencies and types. This article also describes how to make a sentence classifier that uses a neural network created using various libraries created for machine learning in the Python programming language. Easily we can capture the teenager behavior and their conversation regarding suicide and we can stop the propaganda of suicide among minors.



Fig 1: Block Diagram

3.1 MODULES DESCRIPTION

Supervised Classification (Training Dataset)

The data has been divided into two parts i.e., training and testing data in the 70:30 ratios. Learning algorithms have been applied on the training data and based on the learning, predictions are made on the test data set.

Supervised Classification (Test Dataset)

The test dataset is 30% of the total data. Supervised learning algorithms have been applied on the test data and the output obtained is compared with the actual output.

Pandas: pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Numpy: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

Matplotlib: matplotlib.pyplot is a plotting library used for 2D graphics

in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits

Scikitlearn: Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

NLTK: NLTK stands for Natural Language Toolkit. This toolkit is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to it with an appropriate response. Tokenization, Stemming, Lemmatization, Punctuation, Character count, word count are some of these packages.

IV. RESULTS AND DISCUSSION

```
[1] from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.svm import SVC
from sklearn import metrics
import string
from nltk.corpus import stopwords
import matplotlib.pyplot as plt
# import seaborn as sns
%matplotlib inline
```

Fig 2. Results Screenshot

```
[ ] import pandas as pd
df=pd.read_excel('/content/sample_data_1.xlsx')
# df.dropna(inplace=True)
```

1: Sucide and 0: Non Sucidal

▶ df.head()

	title	usertext	y
0	I want to end my life	I'm a loser. I don't want to live.	1
1	Call suicide hotline for help	Don't give up. Try to make change. Call the ho...	0
2	I love this song	It sounds great.	0
3	I will commit suicide tomorrow	I'll never wake up the day after tomorrow.	1
4	I want to end my life	I'm a loser. I don't want to live.	1

Fig 3. Results Screenshot

```
[ ] df.shape
```

```
(96, 3)
```

```
[ ] df.groupby('y').describe()
df['length'] = df['title'].apply(len)
df.head()
# df['length'].plot(bins=50,kind = 'hist')
```

	title	usertext	y	length
0	I want to end my life	I'm a loser. I don't want to live.	1	21
1	Call suicide hotline for help	Don't give up. Try to make change. Call the ho...	0	30
2	I love this song	It sounds great.	0	16
3	I will commit suicide tomorrow	I'll never wake up the day after tomorrow.	1	30
4	I want to end my life	I'm a loser. I don't want to live.	1	21

Fig 4. Results Screenshot

```
[ ] df.groupby('y').describe()
```

	length							
	count	mean	std	min	25%	50%	75%	max
y								
0	48.0	23.0	7.074076	16.0	16.0	23.0	30.0	30.0
1	48.0	25.5	4.547620	21.0	21.0	25.5	30.0	30.0

```
[ ] df['length'] = df['title'].apply(len)
df.head()
```

	title	usertext	y	length
0	I want to end my life	I'm a loser. I don't want to live.	1	21
1	Call suicide hotline for help	Don't give up. Try to make change. Call the ho...	0	30
2	I love this song	It sounds great.	0	16
3	I will commit suicide tomorrow	I'll never wake up the day after tomorrow.	1	30
4	I want to end my life	I'm a loser. I don't want to live.	1	21

Fig 5. Results Screenshot

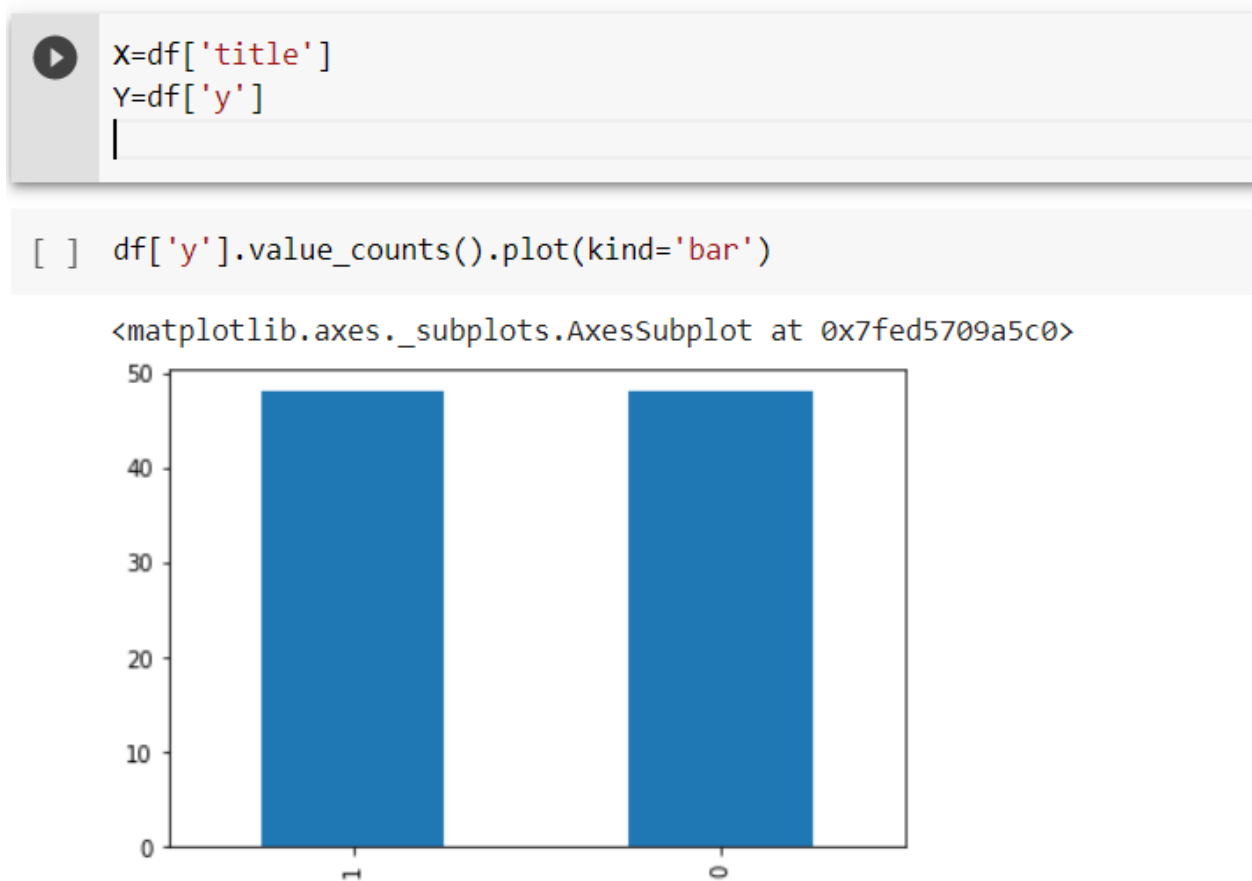


Fig 6. Results Screenshot

```

[ ] count_vect = CountVectorizer()
    count_vect.fit(X)

# transform the training and validation data using count vectorizer object
X_title = count_vect.transform(X)

[ ] from sklearn.model_selection import train_test_split
    X_train, X_test, y_train, y_test = train_test_split(X_title, Y, test_size=0.2)

[ ] print(X_train.shape)
    print(y_train.shape)
    print(X_test.shape)
    print(y_train.shape)

(76, 16)
(76,)
(20, 16)
(76,)

```

Fig 7. Results Screenshot

```

▶ from sklearn.naive_bayes import GaussianNB
nb = MultinomialNB()
nb.fit(X_train, y_train)

preds = nb.predict(X_test)
from sklearn.metrics import confusion_matrix, classification_report
print(confusion_matrix(y_test, preds))
print('\n')
print(classification_report(y_test, preds))
print("Multinomial Accuracy : {}".format(nb.score(X_train, y_train)))

```

```

[[ 8  0]
 [ 0 12]]

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	8
1	1.00	1.00	1.00	12
accuracy			1.00	20
macro avg	1.00	1.00	1.00	20
weighted avg	1.00	1.00	1.00	20

Multinomial Accuracy : 1.0

Fig 8. Results Screenshot

```

▶ title=["Call suicide hotline for help"]

[ ] title=count_vect.transform(title)

[ ] nb.predict(title)

array([0])

```

Fig 9. Results Screenshot

V.CONCLUSION

Among all models for content analysis of the text, this one differs in that it classifies suicidal sentences, which is extremely useful for reducing the death rate from suicide, which exceeds the death rate from

military actions, murders and road accidents. It should be noted that «death groups» were seen not only in «Vkontakte», but also in other social networks and messengers, such as Telegram, Instagram, etc.

VI.FUTURE WORK

This neural network can be used to check the text in social networks to block posts containing suicidal overtones, what will prevent the spread and promotion of suicide among young people who cannot imagine life without social networks. Also, this program can be used for parental control, if the parent does not want to violate the privacy of personal messages of the child, but is worried about him: it is enough to collect data and provide them to the program.

II. REFERENCES

- [1]. Research group «Monitoring of current folklore. Groups of death — from play to moral panic. RANHiGS [RANEPa's typography], 2017. (in Russian)
- [2]. Bryabrina T.V., Gibert A.I., Shtrahova A.V. Experience of content analysis of suicidal statements on the Internet of persons with different levels of suicidal activity. Vestnik YuUrGU [Bulletin of the SUSU], 2016, vol. 9, no. 3, pp. 35–49, DOI: 10.14529/psy160304
- [3]. Raska S. Python and machine learning. DMK Press Publ., 2017, p. 418.
- [4]. Flah P. Machine learning. DMK Press Publ., 2015, p. 400.
- [5]. Liang Wang, Li Cheng, Guoying Zhao. Machine Learning for Human Motion Analysis. IGI Global Publ., 2009, p. 318.
- [6]. Perkins, Jacob. Python Text Processing with NLTK 2.0 Cookbook. — Packt Publishing, 2010.
- [7]. Makinni U. Python for Data Analysis. DMK Press Publ., 2015, p. 482.
- [8]. de Boer, Pieter-Tjerk; Kroese, Dirk P.; Mannor, Shie; Rubinstein, Reuven Y. A Tutorial on the Cross-Entropy Method. Annals of Operations Research, pp. 19–67, DOI:10.1007/s10479-005-5724-z.
- [9]. Jully A., Pal S. Deep learning with Keras. DMK Press Publ., 2017, p. 294.
- [10]. Scholle F. Deep Learning with Python. Piter Publ., 2018, p. 400