Suicidal Tweets Detection in Online social media using Machine Learning

Nithya Bandari, Mounika Kancharla, Umarani Kunsoth

Department of Electronics and Communication Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

ABSTRACT

This project 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.

The study experience of content analysis of suicidal statements on the Internet of persons with different levels of suicidal activity» collects data from the pages of people who have 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. In our system, 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.

Keywords: Suicidal tweets, online social media, machine learning.

1. INTRODUCTION

As per the world Health Organization (WHO), suicide is a primary cause of death among individuals between 15-29 years old across the world. 8, 00,000 of people commit suicide every year leading to increase in suicidal ideation. However, an individual person suicide plays an unsocial act that has overwhelming impact towards relations and families. [1] Several suicidal demises are inevitable and very significant to know the behaviour and the way how individual communicate thoughts and depression for inhibiting such deaths. Suicidal avoidance predominantly focuses on monitoring and observation of suicidal efforts and self-harm tendencies. The existence of content related to suicidal ideation plays a major role on the internet for the people seeking for help and offering support through the younger generation. [2] It is observed that, social media data from different blogs and websites (Facebook, Twitter etc.) are used to recognize the affected individuals instantly to offer help. Suicidal behaviour refers to all promising act of self-harm causing death, while suicidal ideation relates to depressive feeling of planning suicide or killing oneself. Although twitter deliver a chance to know the problem of an individual and to provide a potential way for the intervention of both in social level and individual for suicide prevention there exist no better practices using social media [3]. Suicide avoidance by suicidal identification is a best approach to radically reduce suicidal rates. The major experimental application of this work lies in its flexibility to any web-based social network that should be easily adaptable, wherein it tends to be utilized straightforwardly for breaking down text-based tweets posted by its clients and the tweets are flagged if its contents are related to suicidal thoughts. [4] In recent years, many existing studies focused on n-grams, like 3- grams and 5-grams that are used as keywords and phrases as search terms for suicidal prediction. [5] The objective is, to discover opinion, identify sentiment based on tweets posted by the people and classify them for decision making and suicidal prediction by lexicon and machine learning approach and to identify the suicide prediction level of the twitter users based on the twitter dataset by,

- Creating the dataset to extract knowledge from the patterns in posted tweets showing suicidal tendency by annotated data.
- To increase the performance by using sentiment analysis and classification method for better suicidal prediction.
- Validating the proposed method by comparing with different classifiers.

So, twitter data consisting of raw information is collected from different resources using “Twitter API” and to examine suicidal ideation.

1.1 Problem Statement

Over recent years, social media has become a powerful “window” into the mental health and well-being of its users, mostly young individuals. It offers anonymous participation in different cyber communities to provide a space for a public discussion about socially stigmatized topics. Generally, more than 20% of suicide attempters and 50% of suicide completers leave suicide notes. Thus, any written suicidal sign is viewed as a worrying sign, and an individual should be questioned on the existence of individual thoughts. In addition, social media with its mental health-related forums has become an emerging study area in computational linguistics. It provides a valuable research platform for the development of new technological approaches and improvements which can bring a novelty in suicide detection and further suicide risk prevention. It can serve as a good intervention point.

1.2 Motivation

In this modern age, about half of the world's population dwells on the internet. Teens are engaged very much in a virtual world such as social media like Twitter. This platform is used by teens to share their feelings, sentiments, and even cryptic messages. With the growing issue of suicide at hand and as the social media sites grow in popularity, this platform is increasingly associated with cyber bullying and in some extreme cases teen suicidal. Suicide is the second leading cause of death globally among people who are 15 to 29 years of age.

2. LITERATURE SURVEY

Metzler, et al. [6] proposed detecting potentially harmful and protective suicide-related content on Twitter: machine learning approach. The authors included a majority classifier, an approach based on word frequency (term frequency-inverse document frequency with a linear support vector machine) and 2 state-of-the-art deep learning models (Bidirectional Encoder Representations from Transformers [BERT] and XLNet). The first task classified posts into 6 main content categories, which are particularly relevant for suicide prevention based on previous evidence. These included personal stories of either suicidal ideation and attempts or coping and recovery, calls for action intending to spread either problem awareness or prevention-related information, reporting of suicide cases, and other tweets irrelevant to these 5 categories.

Aldhyani, et al. [7] proposed detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. initially, it is essential to develop a machine learning system for automated early detection of suicidal ideation or any abrupt changes in a user’s behavior by analyzing his or her posts on social media. The authors propose a methodology based on experimental research for building a suicidal ideation detection system using publicly available Reddit datasets, word-embedding approaches, such as TF-IDF and Word2Vec, for text representation, and hybrid deep learning and machine learning algorithms for classification. A convolutional neural network and Bidirectional long short-term memory (CNN–BiLSTM) model and the machine learning XGBoost model were used to classify social posts as suicidal or non-suicidal using textual and LIWC-22-based features by conducting two experiments. To assess the models' performance, the authors used the standard metrics of accuracy, precision, recall, and F1-scores.

Haque, et al. [8] proposed A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. Initially, With the proper exploitation of the information in social media, the complicated early symptoms of suicidal ideations can be discovered and hence, it can save many lives. This study offers a comparative analysis of multiple machine learning and deep learning models to identify suicidal thoughts from the social media platform Twitter. The principal purpose of their research is to achieve better model performance than prior research works to recognize early indications with high accuracy and avoid suicide attempts. They applied text pre-processing and feature extraction approaches such as CountVectorizer and word embedding, and trained several machine learning and deep learning models for such a goal.

Jung, et al. [9] proposed Suicidality detection on social media using metadata and text feature extraction and machine learning. Metadata features were studied in great details to understand their possibility and importance in suicidality detection models. Results showed that posting type (i.e., reply or not) and time-related features such as the month, day of the week, and the time (AM vs. PM) were the most important metadata features in suicidality detection models. Specifically, the probability of a social media post being suicidal is higher if the post is a reply to other users rather than an original tweet. Moreover, tweets created in the afternoon, on Fridays and weekends, and in fall have higher probabilities of being detected as suicidality tweets compared with those created in other times.

Sakib, et al. [10] proposed Analysis of Suicidal Tweets from Twitter Using Ensemble Machine Learning Methods. Initially, the main challenge is to prevent suicidal cases and detect a suicidal note from one's status, or tweet which will help to provide proper mental support to that person. The main motive of the proposed analysis is to anticipate whether a person's tweet contains suicidal ideation or not with the help of machine learning. To attain the objectives, the authors have used an accurate ensemble classifier that can identify content on Twitter that may potentially hint towards suicidal activity. In this research, they have also used several sets of word embedding and tweet features, and they have compared their model among twelve classifiers models.

Chandra, et al. [11] proposed Suicide Ideation Detection in Online Social Networks. Social network services allow its users to stay connected globally, help the content makers to grow their business, etc. However, it also causes some possible risks to susceptible users of these media, for instance, the rapid increase of suicidal ideation in the online social networks. It has been found that many at-risk users use social media to express their feelings before taking more drastic step. Hence, timely identification and detection are considered to be the most efficient approach for suicidal ideation prevention and subsequently suicidal attempts. The authors used a summarized view of different approaches such as machine learning or deep learning approaches to detect suicidal ideation through online social network data for automated detection, is presented.

Mbarek, et al. [12] propose a new method that automatically detects suicidal users through their created profiles in OSNs. Their contribution consists in considering profiles from multiple data-sources and detecting suicidal users based on their available shared content across OSNs. They extract several types of features from the posting content of users to build a complete profile that contribute to high suicidal user prediction. They employ supervised machine learning techniques to distinguish between suicidal and non-suicidal profiles. Their experiments on a dataset, which consists of persons who had died by suicide, demonstrate the feasibility of identifying user profiles from multiple data-sources in revealing suicidal profiles.

3. PROPOSED SYSTEM

This project 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 stochastic gradient descent (SGD) 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.

The study experience of content analysis of suicidal statements on the Internet of persons with different levels of suicidal activity» 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 SGD classifier. In our system, 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.

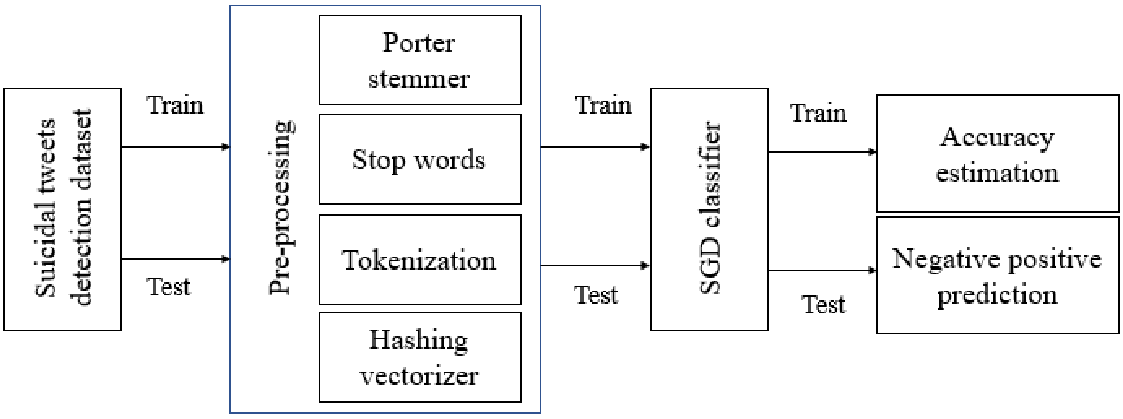


Fig. 1: Proposed block diagram.

3.1 Suicidal tweets data set

2- Columns: Label, Tweet

9120-Rows

3.2 Pre- processing

Data Pre-processing in Machine learning

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

Why do we need Data Pre-processing?

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

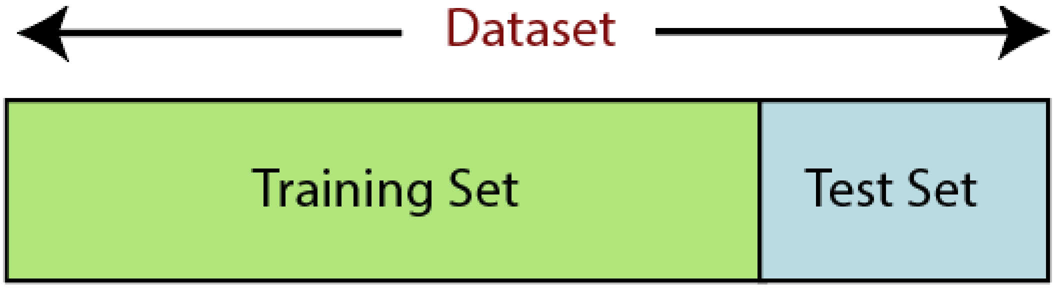
- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

3.2.1 Splitting the Dataset into the Training set and Test set

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model.

Supposeif we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:



Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

3.3 Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is a simple yet very efficient approach to fitting linear classifiers and regressors under convex loss functions such as (linear) support vector machines and logistic regression. Even though SGD has been around in the machine learning community for a long time, it has received a considerable amount of attention just recently in the context of large-scale learning.

SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in text classification and natural language processing. Given that the data is sparse, the classifiers in this module easily scale to problems with more than 10^5 training examples and more than 10^5 features.

Strictly speaking, SGD is merely an optimization technique and does not correspond to a specific family of machine learning models. It is only a way to train a model. Often, an instance of SGDClassifier or SGDRegressor will have an equivalent estimator in the scikit-learn API, potentially using a different optimization technique.

For example, using SGDClassifier(loss='log_loss') results in logistic regression, i.e., a model equivalent to LogisticRegression which is fitted via SGD instead of being fitted by one of the other solvers in LogisticRegression.

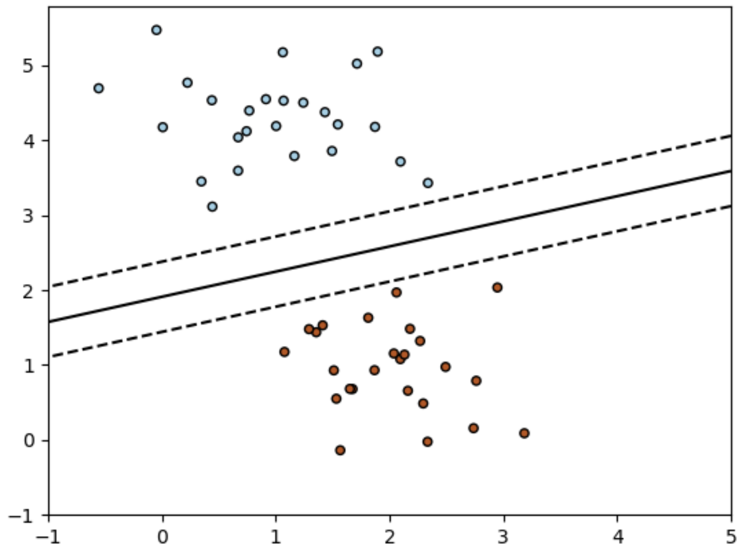
Similarly, SGDRegressor (loss='squared_error', penalty='l2') and Ridge solve the same optimization problem, via different means.

The advantages of Stochastic Gradient Descent are:

- Efficiency.
- Ease of implementation (lots of opportunities for code tuning).

Classification

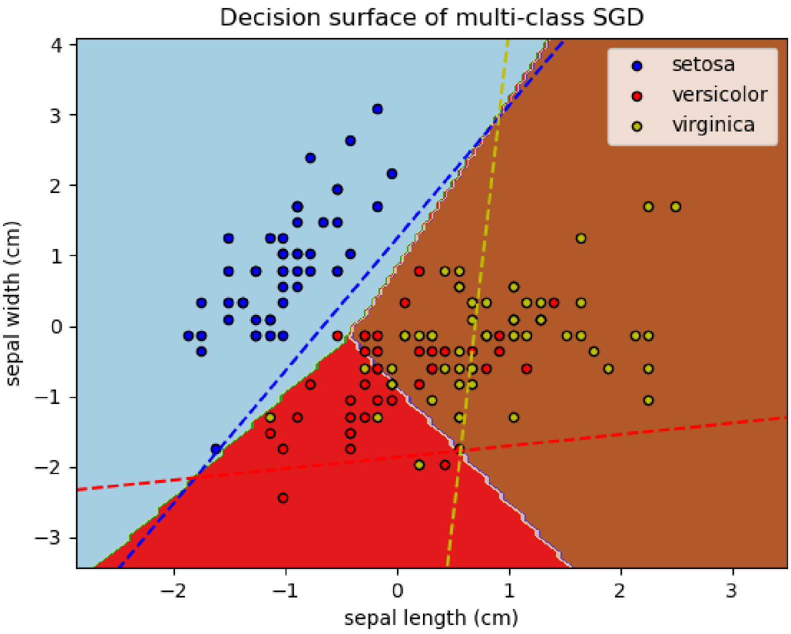
The class SGD Classifier implements a plain stochastic gradient descent learning routine which supports different loss functions and penalties for classification. Below is the decision boundary of a SGD Classifier trained with the hinge loss, equivalent to a linear SVM.



- As other classifiers, SGD must be fitted with two arrays.
- After being fitted, the model can then be used to predict new values.
- SGD fits a linear model to the training data. The coef_ attribute holds the model parameters.
- The intercept_ attribute holds the intercept (aka offset or bias).
- Whether or not the model should use an intercept, i.e., a biased hyperplane, is controlled by the parameter fit_intercept.
- The signed distance to the hyperplane (computed as the dot product between the coefficients and the input sample, plus the intercept) is given by SGDClassifier.decision_function.

SGD Classifier supports multi-class classification by combining multiple binary classifiers in a “one versus all” (OVA) scheme. For each of the K classes, a binary classifier is learned that discriminates between that and all other K−1 classes. At testing time, we compute the confidence score (i.e., the signed distances to the hyperplane) for each classifier and choose the class with the highest confidence. The Figure below illustrates the OVA approach on the iris dataset. The dashed lines represent the three OVA classifiers; the background colors show the decision surface induced by the three classifiers.

In the case of multi-class classification coef_ is a two-dimensional array of shape (n_classes, n_features) and intercept_ is a one-dimensional array of shape (n_classes,). The i-th row of coef_ holds the weight vector of the OVA classifier for the i-th class; classes are indexed in ascending order.



3.4 Porter stemmer

The Porter stemming algorithm (or 'Porter stemmer') is a process for removing the commoner morphological and inflexional endings from words in English. Its main use is as part of a term normalisation process that is usually done when setting up Information Retrieval systems.

3.5 Stop words

When we use the features from a text to model, we will encounter a lot of noise. These are the stop words like the, he, her, etc… which don't help us and just be removed before processing for cleaner processing inside the model. With NLTK we can see all the stop words available in the English language.

3.6 Tokenizer

The breaking down of text into smaller units is called tokens. tokens are a small part of that text. If we have a sentence, the idea is to separate each word and build a vocabulary such that we can represent all words uniquely in a list. Numbers, words, etc. all fall under tokens.

3.7 Hashing vectorizer

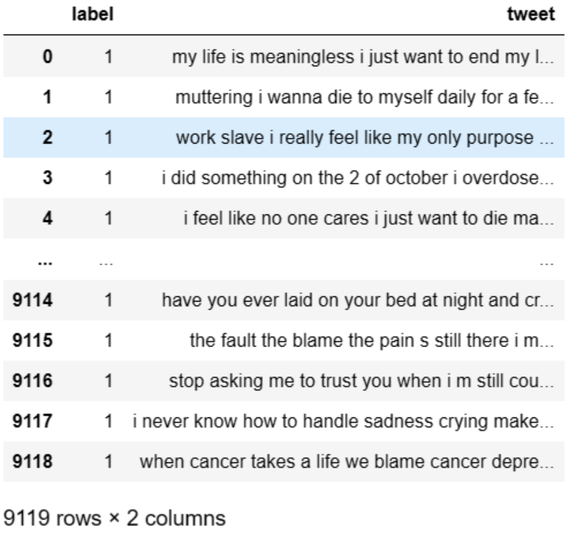
Hashing vectorizer is a vectorizer which uses the hashing trick to find the token string name to feature integer index mapping. Conversion of text documents into matrix is done by this vectorizer where it turns the collection of documents into a sparse matrix which are holding the token occurrence counts.

4. RESULTS AND DISCUSSION

Module description

- Import NLP libraries
- Upload dataset
- Pre-processing
  - ➢ Stemming
  - ➢ Stop words
  - ➢ Tokenizer
- Applying hashing vectorizer
- Building the model (SGD classifier)
- Predicting on test data
- Quality metrics
  - ➢ Accuracy
  - ➢ Probability

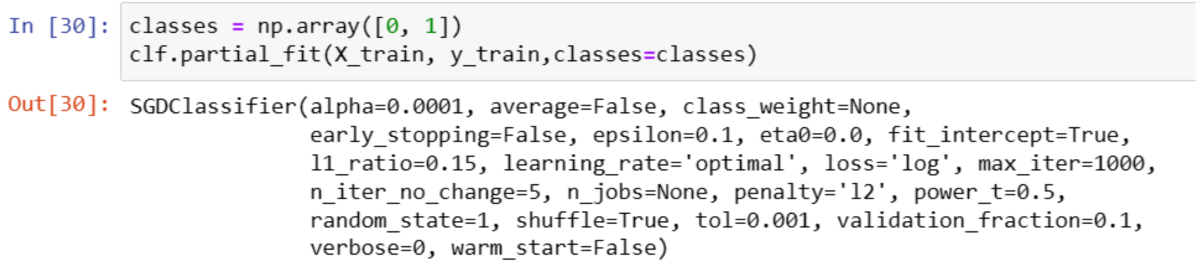
Sample data with labels



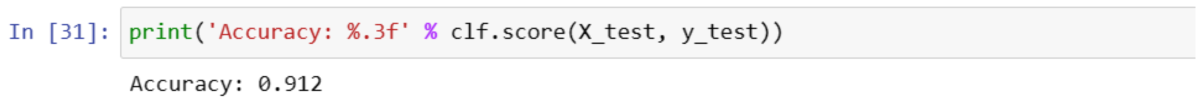
	label	tweet
0	1	my life is meaningless i just want to end my l...
1	1	muttering i wanna die to myself daily for a fe...
2	1	work slave i really feel like my only purpose ...
3	1	i did something on the 2 of october i overdose...
4	1	i feel like no one cares i just want to die ma...
...	...	...
9114	1	have you ever laid on your bed at night and cr...
9115	1	the fault the blame the pain s still there i m...
9116	1	stop asking me to trust you when i m still cou...
9117	1	i never know how to handle sadness crying make...
9118	1	when cancer takes a life we blame cancer depre...

9119 rows × 2 columns

Results with 20% of testing data

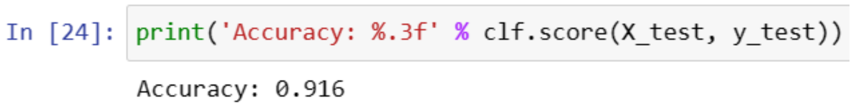


```
In [30]: classes = np.array([0, 1])
         clf.partial_fit(X_train, y_train,classes=classes)

Out[30]: SGDClassifier(alpha=0.0001, average=False, class_weight=None,
                       early_stopping=False, epsilon=0.1, eta0=0.0, fit_intercept=True,
                       l1_ratio=0.15, learning_rate='optimal', loss='log', max_iter=1000,
                       n_iter_no_change=5, n_jobs=None, penalty='l2', power_t=0.5,
                       random_state=1, shuffle=True, tol=0.001, validation_fraction=0.1,
                       verbose=0, warm_start=False)
```



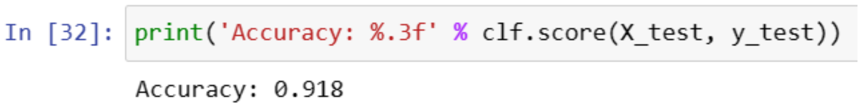
```
In [31]: print('Accuracy: %.3f' % clf.score(X_test, y_test))

         Accuracy: 0.912
```

Results with 25% of testing data

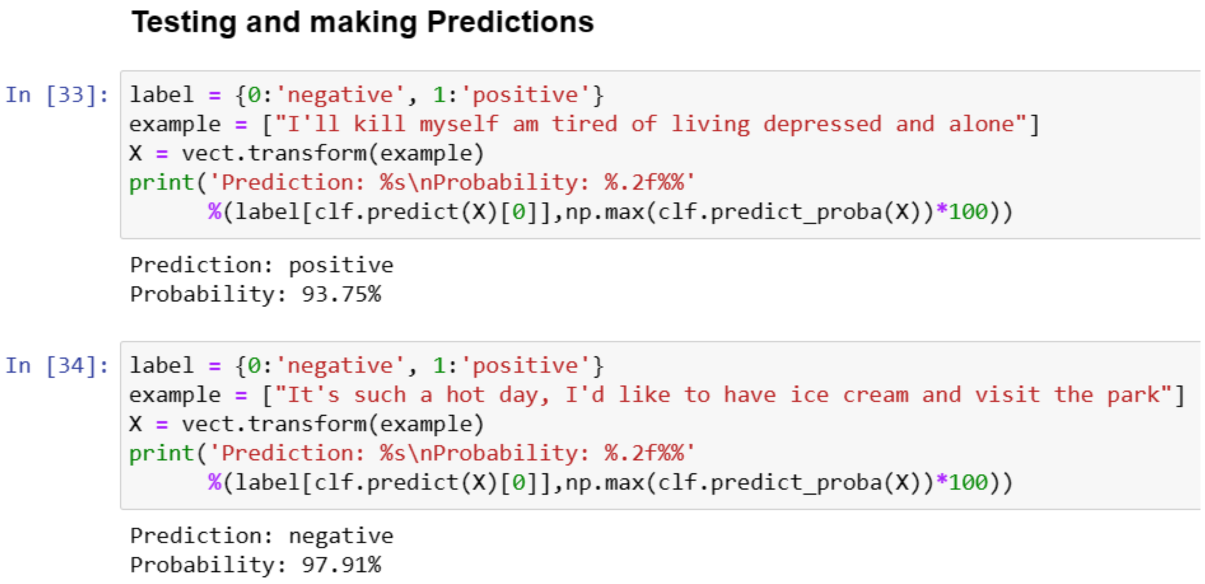


```
In [24]: print('Accuracy: %.3f' % clf.score(X_test, y_test))

         Accuracy: 0.916
```

Results with 40% of testing data



```
In [32]: print('Accuracy: %.3f' % clf.score(X_test, y_test))

         Accuracy: 0.918
```



Testing and making Predictions

```
In [33]: label = {0:'negative', 1:'positive'}
         example = ["I'll kill myself am tired of living depressed and alone"]
         X = vect.transform(example)
         print('Prediction: %s\nProbability: %.2f%%'
               %(label[clf.predict(X)[0]],np.max(clf.predict_proba(X))*100))

         Prediction: positive
         Probability: 93.75%

In [34]: label = {0:'negative', 1:'positive'}
         example = ["It's such a hot day, I'd like to have ice cream and visit the park"]
         X = vect.transform(example)
         print('Prediction: %s\nProbability: %.2f%%'
               %(label[clf.predict(X)[0]],np.max(clf.predict_proba(X))*100))

         Prediction: negative
         Probability: 97.91%
```

5. CONCLUSION

We planned and assessed a novel way to deal with screen the psychological wellness of a client on Twitter. Working off existing examination, we attempted to decipher and evaluate suicide cautioning signs in an online setting (client driven and post-driven social highlights). Specifically, we concentrated on identifying trouble related and suicide-related substance and created two ways to deal with score a tweet: a SGD-based methodology and a progressively conventional machine learning content classifier. To detect changes in enthusiastic prosperity, we considered a Twitter client's action as a surge of perceptions and connected a martingale system to recognize change focuses inside that stream. Our examinations demonstrate that our SGD content scoring approach effectively isolates out tweets displaying trouble related substance and goes about as a ground-breaking contribution to the martingale structure. While the martingale esteems "respond" to changes in online discourse, the change point recognition technique needs enhancement. We could distinguish the genuine change point for one approval case; however, the methodology should be progressively vigorous as for parameter setting and positive changes in speech. setting and positive changes in speech.

Future scope

Future scope will be extended to work on detecting user profiles that are at risk of suicide. We will work on twitter and defined a detection model using a set of rich features including linguistic, emotional, facial, timeline as well as public features to identify twitter profiles. We will use several machine learning methods (mainly classifiers) for the suicidal detection.

REFERENCES

[1] E. Rajesh Kumar, K.V.S.N. Rama Rao, Soumya Ranjan Nayak & Ramesh Chandra (2020) Suicidal ideation prediction in twitter data using machine learning techniques, Journal of Interdisciplinary Mathematics, 23:1, 117-125, DOI: 10.1080/09720502.2020.1721674.

[2] Abdulsalam, Asma, and Areej Alhothali. "Suicidal Ideation Detection on Social Media: A Review of Machine Learning Methods." arXiv preprint arXiv:2201.10515 (2022).

[3] Roy, A., Nikolitch, K., McGinn, R. et al. A machine learning approach predicts future risk to suicidal ideation from social media data. npj Digit. Med. 3, 78 (2020). https://doi.org/10.1038/s41746-020-0287-6.

[4] Rabani, Syed Tanzeel, Qamar Rayees Khan, and A. M. U. D. Khanday. "Detection of suicidal ideation on Twitter using machine learning & ensemble approaches." Baghdad Science Journal 17.4 (2020): 1328-1328.

[5] S. Ji, S. Pan, X. Li, E. Cambria, G. Long and Z. Huang, "Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications," in IEEE Transactions on Computational Social Systems, vol. 8, no. 1, pp. 214-226, Feb. 2021, doi: 10.1109/TCSS.2020.3021467.

[6] Tadesse MM, Lin H, Xu B, Yang L. Detection of Suicide Ideation in Social Media Forums Using Deep Learning. Algorithms. 2020; 13(1):7. https://doi.org/10.3390/a13010007.

[7] Aldhyani THH, Alsubari SN, Alshebami AS, Alkahtani H, Ahmed ZAT. Detecting and Analyzing Suicidal Ideation on social media Using Deep Learning and Machine Learning Models. International Journal of Environmental Research and Public Health. 2022; 19(19):12635. https://doi.org/10.3390/ijerph191912635.

[8] Swain, D., Khandelwal, A., Joshi, C., Gawas, A., Roy, P., Zad, V. (2021). A Suicide Prediction System Based on Twitter Tweets Using Sentiment Analysis and Machine Learning. In: Swain, D., Pattnaik, P.K., Athawale, T. (eds) Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, vol 1311. Springer, Singapore. https://doi.org/10.1007/978-981-33-4859-2_5.

[9] de Carvalho, V.F., Giacon, B., Nascimento, C., Nogueira, B.M. (2020). Machine Learning for Suicidal Ideation Identification on Twitter for the Portuguese Language. In: Cerri, R., Prati, R.C. (eds) Intelligent Systems. BRACIS 2020. Lecture Notes in Computer Science (), vol 12319. Springer, Cham. https://doi.org/10.1007/978-3-030-61377-8_37.

[10] A. Malhotra, R. Jindal, “Deep learning techniques for suicide and depression detection from online social media: A scoping review”, Applied Soft Computing, Vol. 130, 2022, 109713, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2022.109713.

[11] T. H. Sakib, M. Ishak, F. F. Jhumu and M. A. Ali, "Analysis of Suicidal Tweets from Twitter Using Ensemble Machine Learning Methods," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021, pp. 1-7, doi: 10.1109/ACMI53878.2021.9528252.

[12] Chandra, S., Bhattacharya, S., Banerjee (Ghosh), A., Kundu, S. (2021). Suicide Ideation Detection in Online Social Networks: A Comparative Review. In: Mandal, J.K., Mukhopadhyay, S., Unal, A., Sen, S.K. (eds) Proceedings of International Conference on Innovations in Software Architecture and Computational Systems. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore. https://doi.org/10.1007/978-981-16-4301-9_12.