International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SVM Based Model for Detecting Depression Using User Tweets

Prof. Prashant Ahire[1], Kaushal Suryawanshi[2], Rutuja Waykar[3], Sarvesh Saswadkar[4], Ganesh Dhumal[5]

Professor, Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune[1]
Student, Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune[2 3 4 5]

A B S T R A C T

In natural psychological equilibrium, tension may be generally perceived as disturbance. If a per-son is unable to reconcile the expectations imposed on him/ her with his / her capacity to deal to them, so it generates tension and produces burden on mental health. There're many two types of difficulties. Depression may be generally described as psycho-logical equilibrium disruption. One of major research fields of biomedical engineering is Depression detection, as proper Depression prevention could be easy. Several bio signals (i.e., Mri, Rgb, oxygenation, etc.) are available. Which are useful in identifying levels of Depression since these signals indicate distinctive changes in the induction of Depression. In this project, because of the easily accessible recording, we use Twitter tweet as the primary candidate. By modifying the function number and kernel type, multiple SVM model types have been checked.

1. Introduction

Now a day's social media has emerged as a premier method to disseminate information online. Through these online networks, tens of many individuals communicate their thoughts, personal experiences, and social ideals. Therefore, we explore the potential of social media to predict, Major Depressive Disorder in persons. With excess usage of internet, people have started to share their experiences and challenges with mental health disorders through online forums, micro-blogs or tweets. Their online activities inspired many researchers to introduce new forms of potential health care solutions and methods for early depression detection systems.

Detecting emotions by analysing social media's text is a big challenge for researchers. Among all social media, Twitter has become a popular and interesting research area because of its short text. Twitter is a structure that allows users to send and receive short posts called tweets. User can tweet up to 140 characters long and can include links to relevant websites and resources. The goal of depression detection from text is to categorize the actual emotion of the user. It can be taken as a text classification problem. The dataset which we have used in this work contains "depressed" and "non-depressed" polarity of user's tweets. It is a binary classification problem.

Our target work objects to mature and assess different models that classify as Twitter tweets "depressed" or "non-depressed" by using Natural Language Processing, Machine Learning techniques and Deep Learning. Depression is a common mental health disorder and is defined as a single disease with a set of diagnostic criteria. According to the WHO study, there are 322 million people estimated to suffer from depression, equivalent to 4.4% of the global population. Nearly half of the in-risk individuals live in the South-East Asia (27%) and Western Pacific region (27%) including China and India. In most of the countries, depression is still under-diagnosed and left without any adequate treatment which can lead into a serious self-perception and at its worst, to suicide.

This is the main reason they turn to less formal resources such as social media. The reasons that people commit suicide are complicated. People facing depression are highly likely to commit suicide, but many without depression can also have suicidal thoughts. According to the American Foundation for Suicide Prevention (AFSP), it is notice that suicide factors fall under three categories: health factors, environmental factors, and historical.

2. Literature Survey

In this section we will discuss the overview of depression, existing depression detection system by using Different types of machine learning algorithms and will also discuss how SVM is effective in Depression Detection Model. The study on Twitter data in suicide analysis

has become more prevalent as there are known issues with in the psychological assessments that have been studied before. There is a lack of data more importantly textual data for review in analysis of suicide. This review of Twitter offer insight into the day-to-day feelings of individual users [1].

Different algorithms of classification are being used by many researchers to classify the user-generated contents (UGC) from SNS. Aldarwish and Ahmed used various algorithms like Support Vector Machine, Naïve Bayes Classifier. Research done by Hassan et al. shows different machine learning techniques to analyze sentiments and made a comparison of three classifiers: SVM, Naïve Bayes (NB) and Maximum Entropy (ME) regarding the topic. They observed that SVM shows superior results than NB and ME [2]. In sentiment analysis, many machine learning algorithms have been used for classification. Machine learning algorithms, which have shown more successful in classifying the text are NB and SVM. Other machine learning algorithms used for sentiment analysis are centroid classifier, N-gram model, K-Nearest Neighbor.

It has been observed that Support Vector Machine (SVM) can be used for both discrete and continuous data sets also. However, it is mostly used for discrete data sets, i.e., Classification Technique. Rayed Farhad and Farhan Mashuk have used SVM to detect stress using ECG features like QT interval, RR interval and ECG Derived Respiration. They used SVM as there are exactly two class labels in our data, (i.e., stressed and not stressed) [3].

SVM is designed to fundamentally identify and isolate linear separators in search space so as to categorise assorted classes. Ideally, text data is suitable for SVM in view of nature accorded by text. Since, only a selected number of features are immaterial in spite of the tendency to be correlated and organized into linear distinguishing buckets, text data is thus considered as an ideal candidate for SVM. Using SVM, a non-linear decision surface can be made out of original feature space. This can be done by mapping of data instances in a non-linear way to an inner product space where classes can be linearly segregated using hyperplane [4].

3. Methodology

The framework begins with data collection in .csv file. The raw data will be cleaned and pre -processed. The data will then be extracted with respect to the attribute as shown in Fig. i. Next, the data will be fed into classifier which is Support Vector Machine (SVM) and then Final output will showcase whether the person is in depression or not.

3.1 Data Collection
Every machine learning task starts with collection of relevant data from the target source. In this paper Twitter is considered as the target data source for analysis, in the form of User Tweets. This portion covers tasks to compile the training and test datasets. The dataset contains one column named "text" which contains the tweets which were extracted from Twitter.

3.2 Data Processing
The csv file of the data is read and data is pre-processed by some steps. First of all the data is tokenized in which a string is divided into several into several meaningful substring, such as units of words, sentences, or themes. In this case, the first column of the csv file which contains the tweet is extracted and is converted into individual tokens. Tokenisation is followed by stemming which is a step to reduce the words to their root form. This helps us group similar words together after which the commonly used words known as stop words are removed since they are of no use in training and could also lead to erratic results if this type of words is not ignored. Now to improve the quality of the training data, tokenized text is assigned the respective parts of speech to extract only adjectives, nouns and adverbs since other parts of speech are not of much significance. Example: 'Feeling Sad Today' - 'Sad'is extracted, rest are removed.

3.2 Feature Extraction
After Data preprocessing the data extraction or feature extraction is performed to reduce the initial raw data in more manageable groups for processing with respect to the attribues. Feature extraction is the name for process that select variables into features which effectively effectively reduce the amount of data that must be processed, while still accurately and completely describing the original data set.
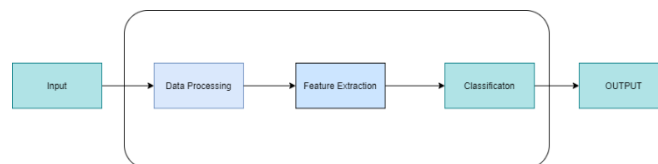

Figure i

3.3 Classification
In this case Support Vector Machine (SVM) is used for classification purpose and also comparison of SVM and NB is tested. As Research done by Hassan et al. shows different machine learning techniques to analyze sentiments and made a comparison of three classifiers: SVM, Naïve Bayes (NB) and Maximum Entropy (ME) regarding the topic. They observed that SVM shows superior results than NB and ME.
Support Vector Machine (SVM) is a supervised learning algorithm SVM analyse the data and recognise the patterns used for the classification. SVM is designed to fundamentally identify and isolate linear separators in search space so as to categorise assorted classes. Ideally, text data is suitable for SVM in view of nature accorded by text. Since, only a selected number of features are immaterial in spite of the tendency to be correlated and organized into linear distinguishing buckets, text data is thus considered as an ideal candidate for SVM. Using SVM, a non-linear decision surface can be made out of original feature space. This can be done by mapping of data instances in a non-linear way to an inner product space where classes can be linearly segregated using hyperplane. SVM classifies the given set as one or the other of two categories. SVM deals both with linear and non-linear classification. it can efficiently perform non-linear classification with kernel trick. It is done by mapping the input set to high dimensional feature space. Polynomial, Gaussian radial basis function (RBF), Laplace RBF kernel, Hyperbolic tangent kernel and Sigmoidal kernel are the types of kernels included in it.

4.   Results

By using machine learning technique like Support Vector Machine (SVM) we are able to detect from the given text that the person is in depression or not as shown in Fig. ii.
Also, by uploading image we can extract text from the image and by using the extracted text it has been detected whether the person is in depression or not as shown in Fig. iii
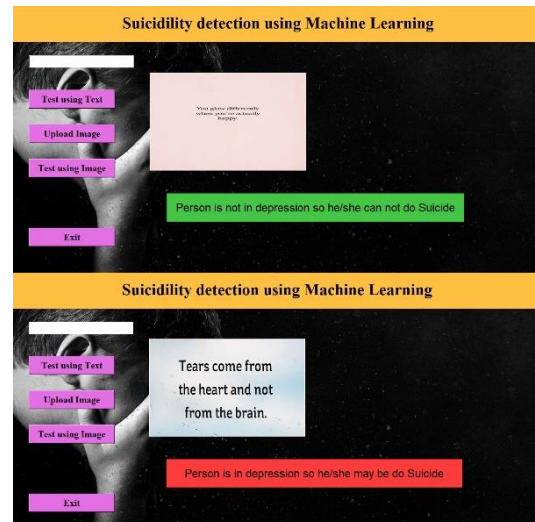

Figure ii

Figure iii

5.   Future Work & Conclusion

Depression is a serious social problem in the world. Efforts have been made in studying the language formation to detect suicidal ideation using open social media data. In the most recent study tweets are manually tagged by human annotators before machine learning algorithms are applied to classify if users are in depression or not, which is not efficient enough to detect suicidal ideation. In this study, we propose a depression detection framework that requires minimum human efforts in annotating data. Support Vector Machine (SVM) has been used here to detect whether the person is in depression or not. However, it is difficult to accurately collect data about the way people feel or think versus how they behave.

Supervised learning classification have some limitations and thus these classifications cannot grant a human level accuracy in prediction of depression through text data. In future, suggestion can be added to the model to reduce number of false positives. This will increase the accuracy of sentiment analysis for depression detection. Many factors are interrelated with suicide, such as mental health, economic rissues, gun prevalence, daylight patterns, divorce laws, media coverage of suicide, and addictions like alcohol use. A better understanding of suicidal intention can definitely provide a guideline for effective detection of depression. A new direction for research is to equip deep learning models with common sense reasoning.

References:

1.   Samah J. Fodeh, Taihua Li, Kevin S. Menczynski, "Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter" in ICDW 2019.

2.   Nafiz Al Asad, Md. Appel Mahmud Pranto, "Depression Detection by Analyzing Social MediaPosts of User" in   2019   IEEE International Conference on Signal Processing, Information, Communication & Systems.

3.   Md Fahim Rizwan, Rayed Farhad, Farhan Mashuk, "Design of a Biosignal Based Stress Detection System Using Machine Learning Techniques" in International Conference on Robotics,Electrical and Signal Processing Techniques, 2019

4.   Raktim Kumar Dey, Debabrata Sarddar, Indranil Sarkar, Rajesh Bose, "A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms" in INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 05, MAY 2020.

5.   Kuhaneswaran A/L Govindasamy, Naveen Palanichamy "Depression Detection Using Machine Learning Techniques on Twitter Data" in IEEE 2021.

6. Shaoxiong Ji, Shirui Pan, *Member, IEEE*, Xue Li, "Suicidal Ideation Detection: A Review of MachineLearning Methods and Applications" in IEEE Transaction on computational social systems.

7. Mandar Deshpande, Vignesh Rao, "Depression Detection using Emotion Artificial Intelligence" in Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017).

8. Sanzana Karim Lora, Nazmus Sakib, Shahana Alam Antora, Nusrat Jahan, "A Comparative Study to Detect Emotions from Tweets Analyzing Machine Learning and DeepLearning Techniques" in International Journal of Applied Information Systems (IJAIS) *Volume 12– No.30, June 2020.*

9. A. Hassan and A. Mahmood, "Deep learning approach for sentiment analysis of short texts," in2017 3rd international conference on control, automation and robotics (IC-CAR), pp. 705–710, IEEE, 2017.

10. A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in2017 7th International Annual Engineering Seminar (InAES), pp. 1–4, IEEE, 2017.