JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Mental illness/Depression Detection from social media data using machine learning techniques

Nirwan University, Jaipur, Computer Science and Engineering
Dr. Pharindra Sharma, Dambarbhadur Rawal(Research Scholar)

Abstract

In the current days maximum people like teenagers, adult are using social media network. One of the biggest differences in the lives of current teenagers and young adults, as compared to earlier generations, is that they spend much time on social media network to connecting directly with peoples and more time connecting electronically, principally through social media. Some experts see the rise in depression or some mental illness as evidence because social media users are sitting on online more than hours. Twitter, Facebook becomes the most popular social media platform that allows people to share the information through small messages called tweets on a real time basis. The proposed method uses concepts like Natural Language Processing for text analysis. This system will analyze the sentiments as positive or negative using Text blob's sentiment method based on that the system will detects the different levels of depression and to find mental illness using machine learning.

Keywords: Machine Learning, Classification Model, Social network, Emotions

Introduction

The proliferations of internet and communication technologies, especially the online social networks have rejuvenated how people interact and communicate with each other electronically. The applications such as Facebook, Twitter, Instagram and alike not only host the written and multimedia contents but also offer their users to express their feelings, emotions and sentiments about a topic, subject or an issue online. On one hand, this is great for users of social networking site to openly and freely contribute and respond to any topic online; on the other hand, it creates opportunities for people working in the health sector to get insight of what might be happening at mental state of someone who reacted to a topic in a specific manner. In order to provide such insight, machine learning techniques could potentially offer some unique features that can assist in examining the unique patterns hidden in online communication and process them to reveal the mental state (such as 'happiness', 'sadness', 'anger', 'anxiety', depression) among social networks' users. Moreover, there is growing body of literature addressing the role of social networks on the structure of social relationships such as breakup relationship, mental illness ('depression', 'anxiety', 'bipolar' etc.), smoking and drinking relapse, sexual harassment and for suicide ideation [1, 2].

In this study, we aim to analyze Facebook data to detect any factors that may reflect the depression of relevant Facebook's users. Various machine learning techniques are employed for such purpose. Considering the key objective of this study, the following are subsequent research challenges addressed in paper.

Define what depression is and what are the common factors contributing toward depression.

What are the factors to look for depression detection in Facebook comments?

How to extract these factors from Facebook comments?

What is the relationship between these factors and attitudes toward depression?

When is the most influential time to communicate within depressive Indicative Facebook user?

What are the most influential machine learning techniques for detection of depression in Facebook comments?

In the context of above mentioned challenges, we analyse depression from Facebook users' data [3, 4]. As users express their feeling as a post or comments in the Facebook platform, sometimes their posts and comments refer to as emotional state such as 'joy', 'sadness', 'fear', 'anger', or 'surprise' [5, 6]. We analyze various features of Facebook comments by collecting data through an effective method of machine learning classification techniques and to make overall judgements regarding their various parts. In this study, we used publically available Facebook data (from bipolar, depression and anxiety Facebook page) containing users' comments. Once we access the data, it was cleaned from any inconsistency and then analyzed by a software application called LIWC [7, 8].

In this study, we examine various linguistic cues which help to detect emotion cause events: the position of cause event and experiencer relative to the emotion keyword: emotional process like positive emotion (e.g. 'happy', 'love', 'nice'), negative emotion (e.g. 'worthless', 'loser', 'hurt', 'ugly', 'nasty'), sadness (e.g. 'worry', 'crying', 'grief', 'sad'), anger (e.g. 'stop', 'shit', 'hate', 'kill', 'annoyed') and anxiety (e.g. 'worried', 'fearful'). A temporal process like present focus (e.g. 'today', 'is', 'now'), past focus (e.g. 'ago', 'did', 'talked') and future focus (e.g. 'shall', 'may', 'will', 'soon'). Linguistic words like articles (e.g. 'a', 'an', 'the'), prepositions (e.g. 'for', 'in', 'of', 'to', 'with', 'above'), auxiliary verbs (e.g. 'do', 'have', 'am', 'will'), conjunctions (e.g. 'and', 'but', 'whereas'), personal pronoun (e.g. 'I', 'them', 'her', 'him'), impersonal pronouns (e.g. 'it', 'it's', 'those'), verbs (e.g. 'go', 'good') and negation (e.g. 'deny', 'dishonest', 'no', 'not', 'never').

Social Networking

The term social media refers to a computer-based technology that facilitates the sharing of ideas, thoughts, and information through virtual networks and communities. People engage with social media via a computer, mobile phone or tablet.The applications such as Facebook, Twitter, Instagram etc.

Purpose

In this study, we aim to perform analysis on social media platform data collected from an online public source. To investigate the study of this data is to find the correct way of preventing from different incident happen in human life. The propose machine learning technique as an efficient and scalable method for this analysis of social media data used by users.

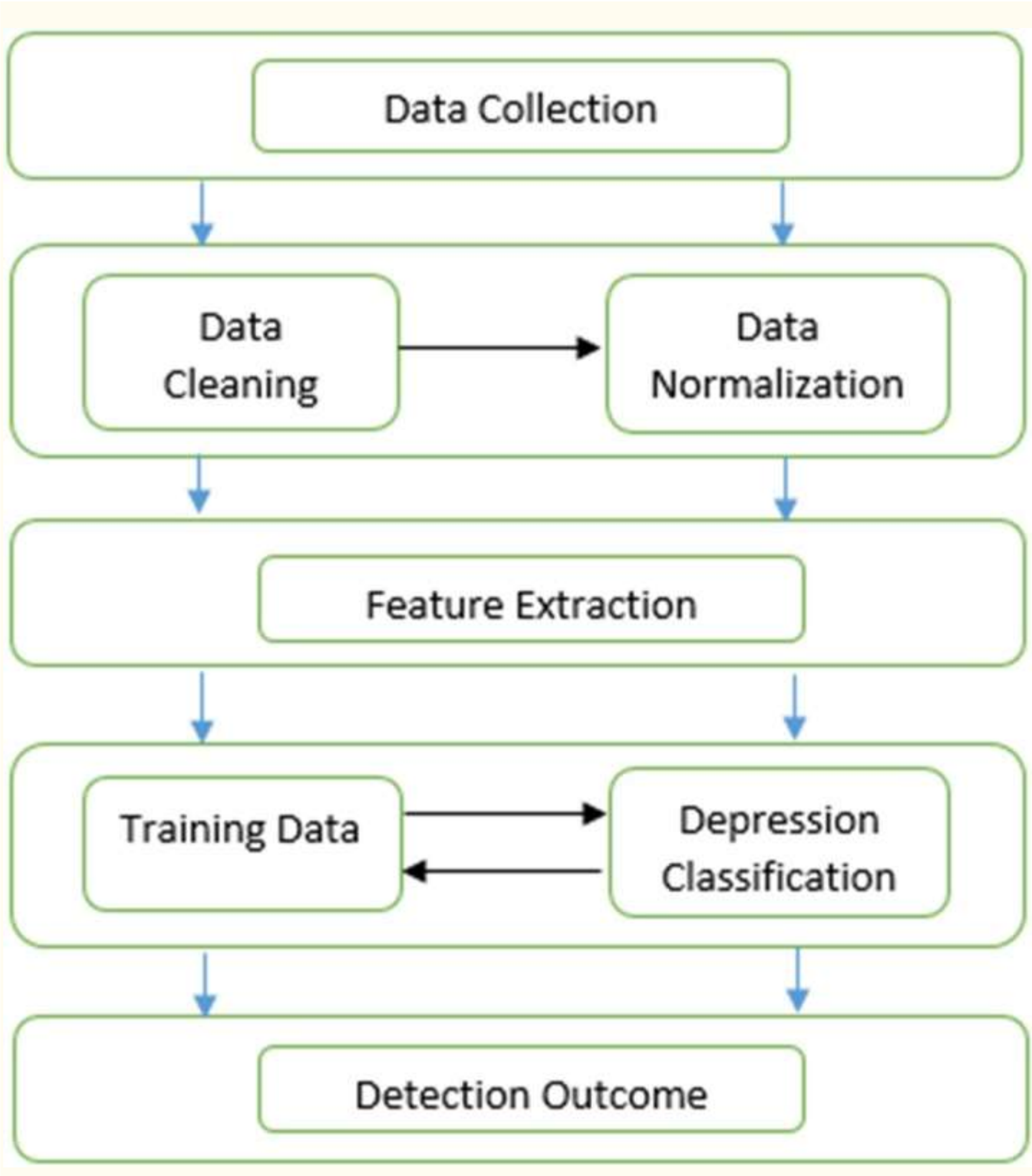
Objective

The main objective of the Study is to build a machine learning model for predict a behavior of person who uses social media platforms. Using this model, it is possible to classify behavior of person who using social media platforms. This CNN based model will help to predict the behavior of person who using social media. There are multiple keywords used in this research analytic tool, python, CNN-model, Data- Augmentation, Data scraping tool etc.

Methods

Although diagnosis of depression using social networks data has picked an established position globally, there are several dimensions that are yet to be detected. In this study, we aim to perform depression analysis on

Social media data collected from an online public source. To investigate the effect of depression detection, we propose machine learning technique as an efficient and scalable method.



Tools & Technique

For analysis the raw data we use NCapture for collecting data from Social Media. After collecting the raw data from Social Media, it was analyzed by using LIWC Software. LIWC is the heart of the text analysis strategy and can process text on a line by line. The analysis is conducted using MATLAB.

Data set exploration

We worked on Facebook users' comments for depressive behavioral exploration and detection. We collected data from the social network . Preparing of social network data, in particular Facebook user's comments is one of the primary challenges which bear information on whether or not they could contain depression bearing content. To tackle this issue we use NCapture for collecting data from Facebook . For qualitative data analysis, NCapture is a powerful tool in the world today. It is intended to enable to arrange, break down and discover knowledge in unstructured data like open-ended survey responses, social media, interviews, articles and web content. Furthermore it gives a place to arrange and deal with material to discover knowledge in a more proficient way.

Feature extraction

To describe and demonstrate amongst depressive and non-depressive posts, we extract the different features in view of psycholinguistic measurements from the user's post. It is clarified briefly as follows:

Psycholinguistic features LIWC is a psycholinguistic vocabulary package made by psychological analysts to perceive the different affective, intellectual, and etymological parts lies on user's verbal or written correspondence. It returns more than 70 different factors with higher level of psycholinguistic features,

Measuring depressive behavior

We presented a set of attributes like emotional process, temporal process, and linguistic style that can be used to characterize the depressive behaviors of users. Our dataset consists of five emotional variables (positive, negative, sad, anger, anxiety), three temporal categories (present focus, past focus and future focus), and 9 standard linguistic dimensions (e.g., articles, prepositions, auxiliary verb, adverbs, conjunctions, pronoun, verbs and negations) [30–36]. We calculate their values by the standard LIWC2015 scales. A complete list of the standard LIWC2015 scales including examples of our dataset is included in Table 4.

Emotional processes Emotion process, a complex experience of consciousness, bodily sensation, and behaviour that reflects the personal significance of a thing, an event, or a state of affairs. The analysis of the emotional comments of social network data can be leveraged to produce reliable predicts in a variety of circumstances [25]. We use psycholinguistic dimensions for considering five features of the emotion state manifested in the comments: positive affect (PA), negative affect (NA), sadness affect (SA), anger affect (AA), and anxiety affect (AnA) [37–41].

Temporal process

Generally, temporal process word provides information about past focus category, present focus category and future focus category of how people are referencing each other and their degree of emotionality.

Linguistic process

Linguistics process is one of the largest parts of LIWC psycholinguistics vocabulary package. It was intended to quantify word use in mentally significant classifications. Also it has been effectively used to recognize connections between people in social co-operations, including relative status, trickiness, and the nature of close relationship. So, In our study we use nine specific linguistics features (articles, prepositions, auxiliary verbs, adverbs, conjunctions, personal pronoun, impersonal pronouns, verbs, and negations) to characterize user comments for our experimental analysis.

Classification model

This stage constructs prediction model for depression post/comments recognition, by considering the psycholinguistic features as input. Considering our training corpus $B = p_1; p_2 \ldots.. p_n$ of n posts/comments, such that each post/comments pi is labeled with the class either as depressive or non-depressive, where $L = l_1|l_2$. The task of a classifier f is to find the corresponding label for each posts/comments.

$$f{:}B{\in}Lf(p)=l$$

In this work, we employ four popular classifiers: Support Vector Machine (SVM), Decision Tree, Ensemble, and k-Nearest Neighbor (kNN).

Support Vector Machines (SVM) Support Vector Machines also known as support vector networks. It is a non-probabilistic linear binary classifier that analyzes data for classification or anomaly detection. It builds a hyperplane into high dimensional feature space and finds a hyperplane that isolates the data into two classes with the biggest separation to the closest training data purpose of any class.

Decision Tree (DT) Decision tree is a simple and all around used classification based systematic approach that makes the hierarchical tree from the training dataset. The state of decision tree is to divide the data hierarchically that have different characteristics. For instance of text documents classification, roots are commonly identified in terms and internal individual nodes may be sub-divided to its children in view of the yes or no of a term in the document.

Ensemble Ensemble methods use multiple learning algorithms of decision tree for better predictive performance.

K-Nearest Neighbor (KNN) K-Nearest Neighbor (KNN) is a non-parametric approach use to discover the distances from point of interest to points in training set.

Experimental analysis

In this study, we examine the execution of various classifiers for depression detection in a shorter time.

Data analysis

The analysis is conducted using MATLAB 2016b. We applied four major classifiers: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision trees (DT), and Ensemble. Each classifier has sub-classifiers such as *Decision trees*—Simple DT, Medium DT, and Complex DT; *SVM*—Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian; *KNN*—Fine, Medium, Coarse, Cosine, Cubic and Weighted, *Ensemble*—Boosted tree, Bagged tree, Subspace discriminant, Subspace KNN, RUSBoosted Tree [42–44].

Using the above classification techniques, we examined detection performance of Facebook user comments. To comprehend the significance of different feature types, we applied four classifiers techniques each utilizing: emotional process, linguistic style, temporal process and all features. The results of the analysis are reported in Tables 5 and and66 that suggests Decision Tree as best performing model. Although KNN gives the high precision but Decision Tree gives the highest result for recall and F-measure relating to the class of depression indicative comments of Facebook user. Similarly, for linguistic style Decision Tree gives the highest result for precision, recall and F-measure.

Conclusion

In this paper we have exhibited the capability of using Social media as a tool for measuring and detecting major depression/mental illness among its users. To give a clear understanding of our work, numbers of research challenges were stated at the start of this paper. The analytics performed on the selected dataset, provide some insight on the research challenges. We find out that the person who using social media network we have to identification depression level of that user.

References

1. Scott J. *Social network analysis.* Thousand Oaks: Sage; 2017. [Google Scholar]

2. Serrat Olivier. *Knowledge Solutions.* Singapore: Springer Singapore; 2017. Social Network Analysis; pp. 39–43. [Google Scholar]

3. Mikal J, Hurst S, Conway M. Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: a qualitative study. In: Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality. 2017.

4. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol.* 2016;**9**:77–82. doi: 10.1016/j.copsyc.2016.01.004. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

5. Ofek Nir, Katz Gilad, Shapira Bracha, Bar-Zev Yedidya. *Advances in Knowledge Discovery and Data Mining.* Cham: Springer International Publishing; 2015. Sentiment Analysis in Transcribed Utterances; pp. 27–38. [Google Scholar]

6. Yang Y, et al. User interest and social influence based emotion prediction for individuals. In: Proceedings of the 21st ACM international conference on Multimedia. 2013. New York: ACM.

7. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol.* 2010;**29**(1):24–54. doi: 10.1177/0261927X09351676. [CrossRef] [Google Scholar]

8. Pennebaker JW, Francis ME, Booth RJ. Linguistic inquiry and word count: LIWC 2001, vol. 71. Mahway: Lawrence Erlbaum Associates; 2001. p. 2001.

9. Holleran SE. *The early detection of depression from social networking sites.* Tucson: The University of Arizona; 2010. [Google Scholar]

10. Greenberg LS. Emotion-focused therapy of depression. *Per Centered Exp Psychother.* 2017;**16**(1):106–117. [Google Scholar]

11. Haberler G. *Prosperity and depression: a theoretical analysis of cyclical movements.* London: Routledge; 2017. [Google Scholar]

12. Guntuku SC, et al. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci.* 2017;**18**:43–49. doi: 10.1016/j.cobeha.2017.07.005. [CrossRef] [Google Scholar]

13. De Choudhury M, et al. Predicting depression via social Media. In: ICWSM, vol. 13. 2013. p. 1–10.

14. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM; 2013.

15. O'Dea B, et al. Detecting suicidality on Twitter. *Internet Interv.* 2015;**2**(2):183–188. doi: 10.1016/j.invent.2015.03.005. [CrossRef] [Google Scholar]

16. Zhang L, et al. Using linguistic features to estimate suicide probability of Chinese microblog users. In: International conference on human centered computing. Berlin: Springer; 2014.

17. Aldarwish MM, Ahmad HF. Predicting depression levels using social media posts. In: 2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS). 2017.

18. Zhou J, et al. Measuring emotion bifurcation points for individuals in social media. In: 2016 49th Hawaii international conference on system sciences (HICSS). 2016. Koloa: IEEE.

19. Wang Xinyu, Zhang Chunhong, Ji Yang, Sun Li, Wu Leijia, Bao Zhana. *Lecture Notes in Computer Science.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network; pp. 201–213. [Google Scholar]

20. Nguyen T, et al. Affective and content analysis of online depression communities. *IEEE Trans Affect Comput.* 2014;**5**(3):217–226. doi: 10.1109/TAFFC.2014.2315623. [CrossRef] [Google Scholar]

21. Park M, McDonald DW, Cha M. Perception differences between the depressed and non-depressed users in Twitter. In: ICWSM, vol. 9. 2013. p. 217–226.

22. Wee J, et al. The influence of depression and personality on social networking. *Comput Hum Behav.* 2017;**74**:45–52. doi: 10.1016/j.chb.2017.04.003. [CrossRef] [Google Scholar]

23. Bachrach Y, et al. Personality and patterns of Facebook usage. In: Proceedings of the 4th annual ACM web science conference. 2012. New York: ACM.

24. Ortigosa A, Martín JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. *Comput Hum Behav.* 2014;**31**:527–541. doi: 10.1016/j.chb.2013.05.024. [CrossRef] [Google Scholar]

25. Shen G, et al. Depression detection via harvesting social media: A multimodal dictionary learning solution. In: Proceeding of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17). 2017. p. 3838–3844.