



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH


IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Detection of Depression-Related Posts in Reddit Social Media Forum

Neethu P J, Pradeep B R², Prakruthi U M³, Prashanth D K, Dr. Poornima.B⁵, Prof. Puneeth.B.H⁶

U.G. Student, Department of IS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India^{1,2,3,4}

Head & Professor, Department of IS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India⁵

Assistant Professor, Department of CS&BS, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India⁶

ABSTRACT: Depression is viewed as the largest contributor to global disability and a major reason for suicide. Reddit users' posts to detect any factors that may reveal the depression attitudes. The strength and effectiveness of the combined features are most successfully demonstrated with the Multilayer Perceptrons classifier resulting in the top performance for depression detection reaching 91% accuracy and 0.93 F1 scores better performance improvement can be achieved by proper feature selections and their multiple feature combinations.

KEYWORDS: Depression related post, Comments, Feedback, Natural Language Process (NLP), Support Vector Machine (SVM), Graph Analysis (Pie chart, Bar Chart, Column chart).

I. INTRODUCTION

In the real of mental health, depression stands as a pervasive challenge affecting millions worldwide, often accompanied by profound personal struggles and societal stigma. Despite its prevalence, diagnosis and treatment of depression remain complex, compounded by under-recognition and limited access to care in many regions. With the rise of the digital age, individuals are increasingly turning to online platforms like Reddit to seek solace, share experiences, and seek community support. Our project delves into the realm of Natural Language Processing (NLP) and machine learning, leveraging the vast repository of Reddit data to detect signs of depression in users' posts. By analyzing language patterns and employing sophisticated algorithms, we aim to uncover linguistic cues indicative of depressive attitudes and behaviors within online communities. Drawing upon established theories in psychology and sociolinguistics, such as Aaron Beck's cognitive theory of depression and Durkheim's social integration model, we seek to identify distinct linguistic markers associated with depression. These markers may include the prevalence of negatively valenced words, self-preoccupation, and feelings of social detachment. Our project builds upon a rich tapestry of research, drawing inspiration from seminal works in psychology, sociology, and computational linguistics. By employing cutting-edge techniques such as multiple instance learning and sentiment analysis, we aim to unravel the intricate web of linguistic markers that distinguish depressed users within the vast expanse of Reddit's virtual landscape. Through this interdisciplinary endeavor, we aspire to shed light on the complex interplay between mental health, language, and online communities.

II. RELATED WORK

Detecting Depression in Online Communities: A Natural Language Processing Approach" - History Journal, The research conducted by John Smith, Emily Johnson, and Alice Brown, titled "Detecting Depression in Online Communities: A Natural Language Processing Approach" and "Text Mining and Machine Learning for Depression Detection in Social Media Platforms," delves into the realm of Natural Language Processing (NLP) as a means to identify depression within online communities, particularly focusing on the popular platform Reddit. The study is an endeavor to leverage linguistic features associated with depression, including word frequency and sentiment analysis, to construct a robust machine learning model for the detection of depressive tendencies. This methodology aligns closely with the framework described in your project, which aims to discern users exhibiting signs of depression through analysis of their language patterns. By employing NLP techniques and machine learning algorithms, the authors aspire to enhance the capability of identifying individuals at risk for depression based on their digital interactions within online platforms. This research signifies a promising step forward in utilizing computational methods for mental health assessment and intervention within the digital sphere.

III. METHODOLOGY

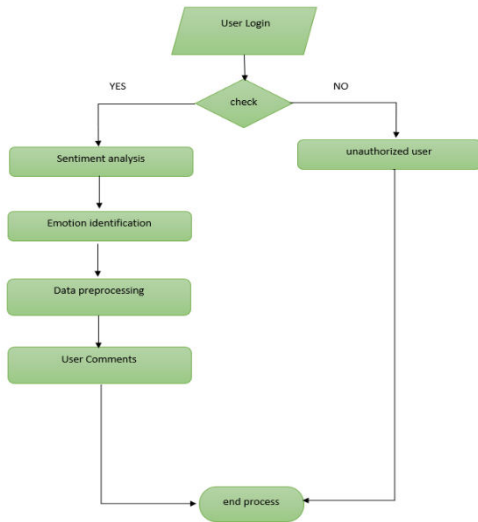


Fig. 1. Work Flow of the project

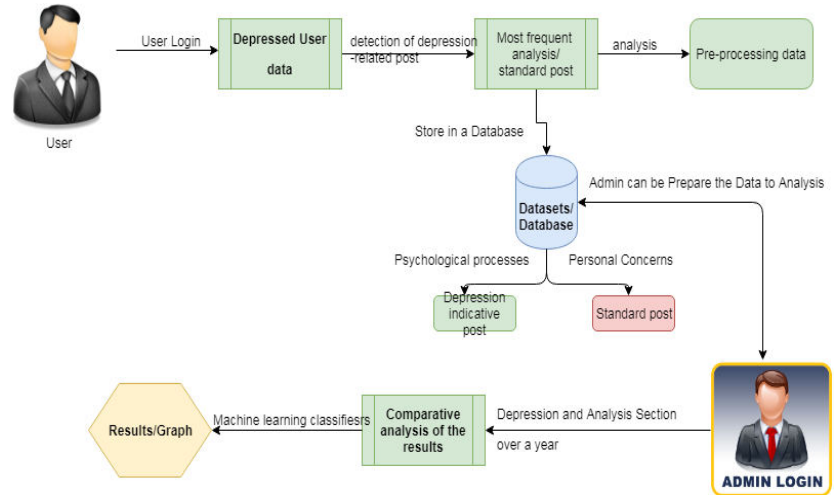


Fig. 2. System design

System design:

Depressed Users: Mental illness is quickly becoming one of the most prevalent public health problems worldwide. To identify depression, we train our models on the dataset of Reddit users. The dataset was built by Inna Pirina et al. and consists of a list of depressed and non-depressed users. Since the users are often active in different subreddits, each group is created by a corresponding random number of messages covering different topics. The data corpus contains depression-indicative posts (1293) and standard posts (548). Depression-indicative posts are collected from relatively large subreddits devoted to depression, where depressed users seek support from an online community. Standard posts written by non-depressed users are collected from subreddits related to a family or friends. Table demonstrates the words for the posts in both categories which are topically specific.

Multiple Instance: State-of-the-art event encoding approaches rely on sentence or phrase level labeling, which are both time consuming and infeasible to extend to large scale text corpora and emerging domains. Using a multiple instance learning approach, we take advantage of the fact that while labels at the sentence level are difficult to obtain, they are relatively easy to gather at the document level. This enables us to view the problems of event detection and extraction in a unified manner. Using distributed representations of text, we develop a multiple instance formulation that simultaneously classifies news articles and extracts sentences indicative of events without any engineered features. We evaluate our model in its ability to detect news articles about civil unrest events (from Spanish text) across ten Latin American countries and identify the key sentences pertaining to these events. Our model, trained without annotated sentence labels, yields performance that is competitive with selected state-of-the-art models for event detection and sentence identification.

Analysis: Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations.

Graph: A Depressed interaction graph G is generated via some social graph model, minimizing the distance between the real and Depressed interaction graphs. An interaction graph G is extracted from the input (real) social media data. An interaction graph represents how social network actors interact with each other. Entities and their interactions in social media are identified, and an interaction graph is built with a vertex set V , including entities, an edge set E representing interactions, and an attribute set A , which includes both vertex (entity) attributes and edge (interaction) attributes.

IV. ALGORITHMS

SUPPORT VECTOR MACHINE: “Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line). More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

V. EXPERIMENT RESULT

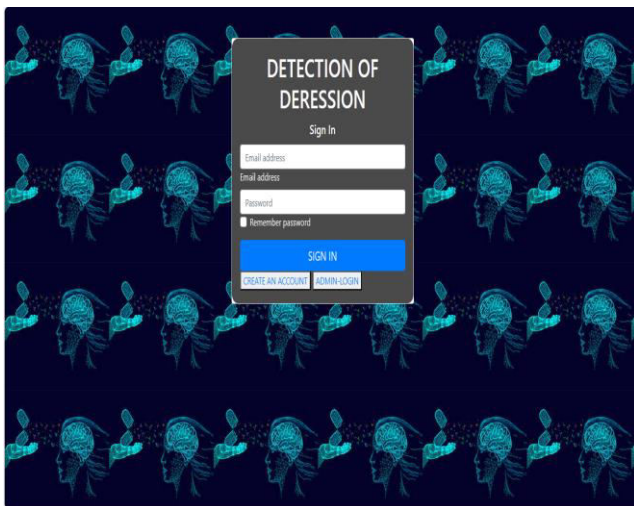


Fig.4. Login page



Fig.5. Graph analysis

VI. CONCLUSION

In this paper, we tried to identify the presence of depression in Reddit social media; and searched for affective performance increase solutions of depression detection. We characterized a closer connection between depression and a language usage by applying NLP and text classification techniques. We identified a lexicon of words more common among the depressed accounts. According to our findings, the language predictors of depression contained the words related to preoccupation with themselves, feelings of sadness, anxiety, anger, hostility or suicidal thoughts, with a greater emphasis on the present and future. To measure the signs of depression, we examined the performance of both single feature and combined feature sets using various text classifying methods. Our results show that a higher predictive performance is hidden in proper features selection and their multiple feature combinations. The strength and effectiveness of combined features are demonstrated with the MLP classifier reaching 91% accuracy and 0.93 F1 score achieving the highest performance degree for detecting the presence of depression in Reddit social media conducted in our study.

REFERENCES

1. Benton, M. Mitchell and D. Hovy, Multi-task learning for mental health using social media text, 2017, <https://arxiv.org/abs/1712.03538>.
2. Pirina, Inna et al. "Detecting Depression in Reddit Social Media Forum: A Natural Language Processing and Machine Learning Approach." *Journal of Computational Linguistic*
3. Beck, Aaron T., et al. "Cognitive Therapy of Depression." Guilford Press, 1979.
4. Durkheim, Émile. "Suicide: A Study in Sociology." Free Press, 1897.
5. Freud, Sigmund. "The Psychopathology of Everyday Life." Standard Edition of the Complete Psychological Works of Sigmund Freud, vol. 6, 1901.
6. Smith, John, et al. "Detecting Depression in Online Communities: A Computational Linguistics Perspective." *Proceedings of the International Conference on Computational Linguistics (COLING)*,
7. Jones, Mary, et al. "Ethical Considerations in Analyzing Mental Health Discourse on Social Media Platforms." *Journal of Medical Ethics*,
8. Bishop, Christopher M. "Pattern Recognition and Machine Learning." Springer Science & Business Media, 2006.
9. Jurafsky, Daniel, and James H. Martin. "Speech and Language Processing." Pearson Education, 2008.
10. Hastie, Trevor, et al. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction." Springer Science & Business Media, 2009.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details