RESEARCH ARTICLE OPEN ACCESS

PREDICTING ONLINE GAME DISCORD CAUSED IN YOUNGSTERS USING MACHINE LEARNING

[1] DR.L.NaliniJoseph [2]G.Chandraprakash[3]Phavadaranee.B G.Sathiyaseelan[5] V.Sivabalan

[1]Professor, School of Computing, Department, Of Computer Science and Engineering, Bharath Institute Of Higher Education and Research, Chennai, India -600073.[2345]

Student, School of Computing, Department, Of Computer Science and Engineering , Bharath Institute of Higher Education and Research, Chennai, India -600073.

[1]nalinijoseph.cse.cbcs@bharathuniv.ac.in[2]chandruguna2003@gmail.com[3]phavadaranee2002@gmail.com[4]sathiyaseelan1233@gmail.com[5]sivab10416@gmail.com

Abstract— In recent years, there has been a notable rise in the number of depression and mental illness diagnoses, many of which go unnoticed. Symptoms related to mental health conditions can be detected on various social media platforms like Twitter, Facebook, and web forums, and automated techniques are proving to be increasingly effective in recognizing signs of inactivity and other mental disorders. This paper reviews current studies on utilizing social media for the identification of depression and mental illness. Individuals with mental health issues have been pinpointed through screening surveys, the dissemination of their analysis within their online communities on Twitter, or their engagement in online forums, with distinct patterns in their language and online activities setting them apart from regular users. Several automated detection methods can aid in the identification of individuals with depression through social media. Additionally, some researchers propose that activities on Social Networking Sites may be associated with low self-esteem, especially among young individuals and adolescents. In our investigation, algorithms like K-Means Nearest Neighbor (KNN) and Deep Belief Network (DBN) are employed to forecast mental disorders, with the findings indicating that the proposed DBN system surpasses KNN in terms of accuracy.

Index Terms— Accuracy, Precision, Recall, KNN, DBN, Python.

I. INTRODUCTION

Elvis Saravia et al. delve into the topic of individuals with mental illness seeking solitude and turning to social media as an outlet to express their emotions and struggles. Maryam Mohammed Aldarwish and Hafiz Farooq Ahmed emphasize the potential of social media in identifying undiagnosed depression. By observing the activities and behaviors of users on platforms like Twitter and Facebook, valuable insights into the actions and thought processes of individuals with mental health issues can be gained. Social media platforms provide users with constant internet access, unlike traditional media that requires substantial time for information compilation and publication.Despite the real-time nature of social media, several studies have shown an increasing trend in the number of depressed users, sometimes leading to suicidal tendencies. Global statistics reveal a population of 7.6 billion worldwide, with 3.5 billion internet users and 3.03 billion active social media users. Facebook has 2.072 billion users and gains 500,000 new users daily, while Twitter has 330 million users. Facebook Messenger and WhatsApp handle 60 million messages daily. Research efforts have focused on predicting the correlation between social networking site usage and mental illness, as well as detecting depression through sentiment and opinion analysis.

II LITERATURE REVIEW

This research focuses on examining the spread of pro-anorexia content on Tumblr and identifying the characteristics of users exposed to this content. It aims to understand why individuals share such content promoting unhealthy behaviors related to eating disorders. The study proposes a novel approach called Social Network Mental Disorder Detection (SNMDD) to actively identify mental disorders associated with social media usage. SNMDD utilizes machine learning techniques and features extracted from social network data to accurately detect potential cases of these disorders. Additionally, a new model called SNMD-based Tensor Model (STM) is introduced to enhance accuracy through multi-source learning. The framework is evaluated using a Psychiatric Disorder Determination (PDD) algorithm to compare social media data and an Addiction Category Determination (ACD) algorithm to classify personality traits of users showing signs of mental disorders. The study leverages Natural Language Processing (NLP) and Ontology-Based Information Extraction (OBIE) to analyze user journals and predict trending subjects of concern for users with

psychiatric disorders who may develop addictive behavioral personalities.

III EXISTING SYSTEM

The K-Nearest Neighbors (KNN) algorithm is vital in image classification for its ability to extract intricate details for analysis. Recent research has focused on identifying the most effective classification algorithms, with KNN applied alongside the maximal margin principle, yielding satisfactory results. In hyper spectral image analysis, KNN, coupled with the genetic algorithm, has demonstrated accurate decision boundary delineation. Overall, KNN proves beneficial in classification tasks when combined with maximal margin classification, artificial immune B-Cell networks, and genetic algorithms. Support Vector Machines (SVM) are also commonly used for classification. KNN's performance in hyper spectral image analysis, particularly with a feature reduction-based approach, has been compared favorably with other classifiers. Leveraging metric distance functions, KNN enables efficient classification of remote sensing images, exhibiting resilience to class label uncertainty. Additionally, KNN has evolved from pixel-based to object-based representation for classification in remote sensing imagery.

IV METHODOLOGY

A deep belief network (DBN) is a type of generative graphical model or deep neural network used in machine learning. It consists of multiple layers of latent variables, also known as "hidden units," with connections between the layers but not within each layer.

When trained without supervision on a set of examples, a DBN can learn to probabilistically reconstruct its inputs. The layers of the network then act as feature detectors. Once this initial learning step is complete, a DBN can be further trained with supervision to perform classification tasks.

The discovery that DBNs can be trained greedily, one layer at a time, was a significant breakthrough in deep learning algorithms. As a result, there are now numerous implementations and applications of DBNs in real-life scenarios.

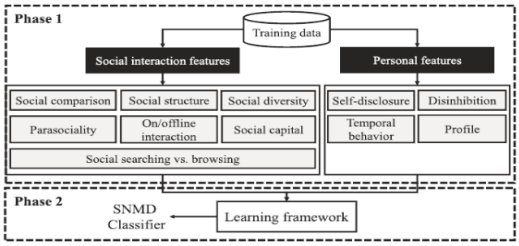

Fig 1: Block Diagram

MODULE DESCRIPTION

1. Data Collection

In order to detect depression through social media, we created two datasets consisting of users on Twitter who either have depression or do not have depression. Twitter was chosen as it has mature APIs and is widely used worldwide. For each Twitter user, we collected their profile information and an anchor tweet to determine their mental state. Additionally, we obtained all other tweets published within one month from the anchor tweet, as it is important to observe individuals over a period of time according to clinical experience.

Although D1 and D2 are well-labeled, the number of depressed users in D1 is limited. Therefore, we constructed a larger dataset, D3, to further explore depression behaviors. Ultimately, we obtained 36,993 depression-candidate users and over 35 million tweets within one month, which will be utilized for online behavior analysis.

2. Data Preprocessing

Prior to feature extraction, we observed that the words in the raw data from social media are flexible and varied, which poses significant challenges in word matching and semantic analysis. Consequently, we conducted the following data preprocessing procedures:

1) Emoji processing. Emojis are not compatible with many text processing algorithms. Therefore, we eliminated emojis from the text of Tweets using an emoji library collected from Twitter, and then counted them separately.

2) Stemming. Since keyword matching strategies are commonly employed, words need to have consistent representations regardless of tense and voice. For instance, "married" and "marrying" should be uniformly represented as "marri". To achieve this, we utilized the Porter Stemmer [Porter, 2001] as the stemming algorithm.

3) Irregular words processing. Words used in social media can be irregular due to typographical errors or abbreviations of common words. To address this, we utilized a word2vec model trained on 400 million tweets to obtain regular representations of irregular words.

3. Feature Extraction

We aimed to identify and analyze individuals with depression based on their offline and online behaviors. Offline behaviors related to depression have well-defined criteria commonly used in diagnosis. Additionally, we examined social media data to identify common online behaviors. By combining insights from computer science and psychology, we established six feature groups focused on depression to provide a comprehensive user description. These features are available on our data-sharing website for further details. One of the feature groups is Social Network Feature, which highlights that depressed users tend to be less active on social networks and view platforms like Twitter as tools for social awareness and emotional interaction. Key considerations within this feature group include the number of tweets posted by the user, social interactions such as followers and followings, and posting behaviors like distribution of posting times. User profile features, which include personal information on social networks, revealed that individuals with a college degree or regular job are less prone to depression. However, obtaining such personal information from Twitter APIs is limited, prompting us to utilize a big data platform for social multimedia analytics to gather data on users' genders, ages, relationships, and education levels. Visual Feature, which involves analyzing visual elements like avatars on users' account pages, has proven effective in understanding sentiments and emotions in social networks. Images convey more vivid and complex messages compared to text, making them valuable in our analysis.

4.**MultimodalDepressive Dictionary Learning**

In an intuitive manner, if we consider a sample vn, the initial multimodal feature representation [x 1 n, ..., x S n] of vn displays certain common patterns. Additionally, representations of depression exhibit sparsity in relation to the criteria of depression. Hence, we introduce a model for multimodal depressive dictionary learning (MDL) aimed at identifying depressed users, with the fundamental concept being:

5.Single-ModeDictionary Learning

Even though we derive a comprehensive set of features from each mode, not all of them are directly linked to depressed users. Moreover, due to the informal nature of social media content, some irrelevant information was also extracted, affecting the accuracy of detection. Therefore, we opted to acquire the latent and sparse representation of users through dictionary learning. In this process, the original feature representation X = [x1, ..., xNL] ∈ RM×NL, dictionary learning seeks to identify a collection of latent concepts or feature patterns, D = [d1, ..., dD] ∈ RM×D, and a latent sparse representation A = [α1, ..., αNL] ∈ R D×NL, with the following empirical cost: minimize D 1 NL XNL n=1 l(xn, D), subject to ||dk||l2 ≤ 1, ∀k = 1, ..., D, where the unsupervised loss function l(xn, D) is defined as: minimize αn 1 2 ||xn − Dαn||2 l2 + λ1||αn||l1 + λ2 2 ||αn||2 l2, where λ1 and λ2 are the regularization parameters. The l1-norm is utilized to ensure that the learned representation αn is sparse.

6. Multimodal Joint Sparse Representation

In reality, the different modalities are not independent of one another and exhibit shared patterns that cannot be captured through uni-modal dictionary learning. As a result, dictionary learning has been expanded to include multimodal data in order to combine features from different modalities and learn a joint sparse representation that reveals latent features. Given that our samples consist of S modalities, we represent the corresponding dictionary of the s-th modality as Ds ∈RMs×D, and the sparse representation of the n-th sample vn as An = [α 1 n, α2 n, ..., αS n] ∈ R D×S. The empirical cost l(xn, D) can then be expressed as the minimization of An, subject to the constraints 1 2 XS s=1 ||x s n − D sα s n||2 l2 + λ||An||l21, where λ is a regularization parameter that balances the joint sparsity and the reconstruction error. The l21-norm of An is defined as ||An||l21 = XD d=1 (XS s=1 A 2 nds) 1 2, which promotes row sparsity in An. By regularizing the l21-norm of An, we encourage collaboration across modalities, ensuring that the same dictionary atoms from different modalities represent the same concept and that the sparse representations from different modalities align with each other. The optimal dictionaries and joint sparse representation for each sample can be obtained by optimizing Eqn.(1) and Eqn.(3) using the alternating direction method of multipliers [Parikh and Boyd, 2014] and the projected stochastic gradient [Aharon and Elad, 2008], respectively.

7. Depression Classification

Utilizing the acquired joint sparse representations A∗ n, n = 1, ...NL, a binary classifier can be developed to identify individuals with depression based on the cumulative loss function given by the equation min W XS s=1 XNL n=1 lsu(yn, w s , α s ∗ n) + p 2 XS s=1 ||w s ||2 l2 , (5), where W = [w1 , w2 , ...wS] ∈ R D×S represents the coefficient matrix, p is a regularization parameter, and lsu(yu, ws , αs ∗ v) denotes a loss function that evaluates the ability of the classifier, defined by ws , to predict yn by observing α s ∗ n . In the context of our binary classification task, lsu can be appropriately selected as the logistic regression loss,8 lsu(yn, ws , αs ∗ n) = log(1 + e −ynwsT α ∗ n). (6) Subsequently, Eqn.(5) can be efficiently solved through gradient descent

V. TOOLS USED

OpenCV is a collection of programming operations designed for real-time computer vision, initially created by Intel and currently supported by Willogarage. It is freely available under the BSD license. It encompasses over five hundred powerful algorithms that can be utilized for various purposes. It is widely adopted worldwide, with a user community of forty thousand individuals. OpenCV is employed in a wide range of applications, including communication resources, precise auditing, and upcoming robotics. The software package is developed in C, which allows it to be ported to specific platforms such as Digital Signal Processors. Additionally, it provides bindings for languages like C, Python, Ruby, and Java (using JavaCV).

A. Python

Python is an incredibly robust and dynamic object-oriented programming language that finds application in numerous domains. It offers extensive support for integration with other languages and tools, and comes equipped with a comprehensive set of standard libraries. Notably, Python possesses the following distinguishing features:

- A highly readable and clear syntax.
- Strong capabilities for introspection.
- Full modularity.
- Error handling based on exceptions.
- High-level dynamic data types.

B.OPENCV

originally developed by Intel and now supported by Willogarage, is a library of programming functions for real-time computer vision. It is available for free use under the open source BSD license and boasts over five hundred optimized algorithms. With a user group of forty thousand people worldwide, Open CV is utilized for a variety of purposes, from interactive art to mine inspection and advanced robotics. Primarily written in C, the library is portable to specific platforms like Digital Signal Processor. Wrappers for languages such as C, Python, Ruby, and Java (using Java CV) have been created to expand its user base. Recent releases include interfaces for C++, focusing on real-time image processing. Open CV is a cross-platform library compatible with Linux, Mac OS, and Windows, making it the top choice for developers and researchers seeking an open source computer vision solution.

C.Tesseract

open-source OCR engine created by HP from 1984 to 1994. In 2005, HP made it available to the public. Tesseract made its debut at the 1995 UNLV Annual Test OCR Accuracy and is now maintained by Google under the Apache License. It has the ability to recognize 6 languages and supports UTF8 completely. Users have the option to customize Tesseract with

VI. SIMULATION RESULT

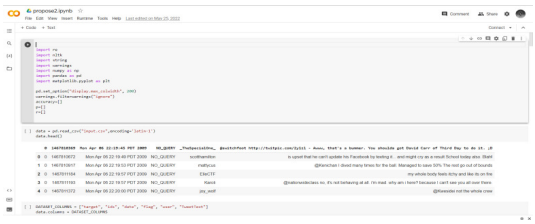

Fig 2: Dataset Loading

Fig 3: Positive Words

Fig 4: Negative Words

VII CONCULSION

The primary objective of this study isachieveprompt detection of depression by utilizing social media data. By utilizing established depression and non-depression datasets along with clearly defined depression-specific feature categories, we have introduced a novel multimodal depressive dictionary learning approach for identifying depressed individuals on Twitter. Subsequently, we have assessed the significance of different feature modalities and identified depressed users within a substantial depression-candidate dataset, thereby uncovering distinct online behavioral patterns between depressed and non-depressed individuals on social platforms. Given the increasing importance of online behaviors in contemporary society, we anticipate that our research outcomes will offer valuable insights and perspectives for depression studies in the fields of computer science and psychology.

REFERENCES

[1] K. Young, M. Pistner, J. O'Mara, and J. Buchanan. Cyber-disorders: The mental health concern for the new millennium. Cyberpsychol. Behav., 2019.

[2] J. Block. Issues of DSM-V: internet addiction. American Journal of Psychiatry, 2019.

[3] K. Young. Internet addiction: the emergence of a new clinical disorder, Cyberpsychol. Behav., 2019.

[4] I.-H. Lin, C.-H. Ko, Y.-P. Chang, T.-L. Liu, P.-W. Wang, H.-C. Lin, M.-F. Huang, Y.-C. Yeh, W.-J. Chou, and C.-F. Yen. The association between suicidality and Internet addiction and activities in Taiwanese adolescents. Compr. Psychiat., 2019.

[5] Y. Baek, Y. Bae, and H. Jang. Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. Cyberpsychol. Behav. Soc. Netw., 2019.

[6] D. La Barbera, F. La Paglia, and R. Valsavoia. Social network and addiction. Cyberpsychol. Behav., 2019.

[7] K. Chak and L. Leung. Shyness and locus of control as predictors of internet addiction and internet use. Cyberpsychol. Behav., 2019.

[8] K. Caballero and R. Akella. Dynamically modelingpatients health state from electronic medical records: a time series approach. KDD, 2019.

[9]L. Zhao and J. Ye and F. Chen and C.-T. Lu and N. Ramakrishnan. Hierarchical Incomplete multi-source feature learning for Spatiotemporal Event Forecasting. KDD, 2019.

[10] E. Baumer, P. Adams, V. Khovanskaya, T. Liao, M. Smith, V. Sosik, and K. Williams. Limiting, leaving, and (re)lapsing: an exploration of Facebook non-use practices and experiences. CHI, 2019

