SEYBOLD Report
ISSN: 1533 - 9211
https://seyboldreport.net/

AN IN-DEPTH ANALYSIS OF DETECTING MENTAL HEALTH DISORDERS THROUGH SOCIAL MEDIA DATA MINING

Dr. Rajendra Singh Kushwah[1]
Research Guide, Dept. of Computer Application, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal-Indore Road, MadhyaPradesh, India

Shah Riteshkumar Rameshchandra [2]
Research Scholar, Dept. of Computer Application, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal-Indore Road, MadhyaPradesh, India

Dr. M V Narayana[3]
Research Co- Guide, Dept. of Computer Science, Guru Nanak Institutions Technical Campus, Khanapur Village, Manchal, Ibrahimpatnam, Telangana 501506

ABSTRACT
The rapid growth of social media platforms has raised concerns about their impact on mental health, leading to the development of techniques for detecting psychological disorders through data mining. This paper proposes a novel system for detecting mental health disorders, particularly in online social network users. Using machine learning techniques, such as Support Vector Machines (SVM), the system is designed to identify at-risk users by analyzing their social media behavior. Key features, including parasocial interactions, online vs. offline communication ratios, and temporal behaviors, are extracted to assess potential mental health issues. The proposed model is evaluated using real datasets from social media platforms like Facebook and Instagram, demonstrating high accuracy in identifying users exhibiting signs of cyber-relationship addiction, net compulsion, and information overload. The results suggest that integrating multiple behavioral and personal features can significantly improve the precision of mental health disorder detection.
Keywords: Social media mining, Psychological disorder detection, Machine learning, Online social networks, Support Vector Machines (SVM)

INTRODUCTION

1. Mental Health

Mental health is the most fundamental and inseparable component of overall health. It is a state in which individuals can experience joy in life while functioning productively, engaging in meaningful relationships, and coping with adversity without losing their ability to function physically, mentally, or socially. Mental health is now recognized as a crucial part of a person's overall well-being, contributing significantly to both physical health and social efficacy. Unfortunately, mental health and mental disorders have not always been given the same priority as physical health and are often neglected.

Mental disorders are highly prevalent, affecting nearly half of the world's population in some way, impacting self-esteem, relationships, and daily functioning. Without proper care, mental health issues can lead to complications such as substance abuse and functional impairments. Maintaining good mental health is essential for leading a fulfilling and healthy life. The integration of a sound mind and controlled emotions is now considered part of modern health definitions, expanding beyond mere physical health to include mental well-being.

Mental health is not a static condition, as it is influenced by a combination of biological, psychological, and social factors. According to the World Health Organization (WHO), mental health is defined as the ability to form harmonious relationships with others and contribute meaningfully to one's community. Additionally, emotional stability and personal satisfaction are vital characteristics of a mentally healthy person. Therefore, good mental health involves adapting to life's demands while maintaining emotional balance and personal contentment.

2. Social Networks

Social networks have emerged as a new global phenomenon, connecting people across the world and facilitating communication, information sharing, and collaboration. They are defined as social structures consisting of individuals or groups connected by shared relationships or interests. With the advent of online social networks (OSNs), the internet has revolutionized the way people interact, providing platforms where users can create profiles, manage their identities, and engage with others.

The rapid development of communication technologies has made it easier for people to build virtual communities, stay connected, and maintain a digital presence. Social networks allow individuals to share content, exchange ideas, and influence others' opinions. This has given rise to a new level of social interaction, where personal connections and group affiliations can be expanded beyond geographical boundaries. The history of social networks traces back to early human interactions around campfires and painted cave walls, where storytelling and shared experiences were central to social cohesion.

The rise of online social networks like Facebook, Twitter, and Instagram has transformed how people communicate, with social network analysis (SNA) becoming a critical tool for understanding the structure and dynamics of these networks. SNA focuses on analyzing relationships and behaviors within networks, providing valuable insights into group dynamics, consumer behavior, and even academic collaborations. The growth of social networks has created an expansive digital ecosystem, where both personal and professional interactions take place on a global scale.

3. Online Social Media

Online social media platforms like Twitter, Facebook, and Instagram have fundamentally changed how people share information and interact with each other. These platforms enable users to express their opinions, share their thoughts, and connect with a global audience in real time. Social media monitoring has become increasingly important, as it provides insights into public opinion, popular trends, and emerging societal issues.

The sheer volume of content generated on social media offers a rich source of data for analysis. Social media data is unstructured and constantly evolving, making it a valuable resource for understanding public sentiment and behaviors. Platforms like Twitter are particularly useful for tracking public opinion on events such as elections, government policies, and consumer preferences. Researchers and organizations use sentiment analysis techniques to categorize and evaluate social media content, identifying trends and forecasting outcomes based on the collective opinions of users.

Despite its many benefits, social media also presents challenges in terms of information overload, misinformation, and the potential for addiction. The constant stream of updates, notifications, and interactions can overwhelm users, leading to decreased productivity and mental health issues such as anxiety and depression. Online social media has therefore become a double-edged sword, offering both opportunities for connection and the risk of negative consequences for individuals' mental and emotional well-being.

4. Information Mining from Social Media

Information mining from social media involves extracting meaningful data from user-generated content on platforms like Twitter, Facebook, and blogs. This process relies on analyzing vast amounts of unstructured data to gain insights into public sentiment, preferences, and behaviors. The interactions on social media provide valuable information that can be used for business intelligence, market analysis, and social research. By processing this data, organizations can better understand their audience and respond to their needs.

Social media mining uses techniques from natural language processing (NLP), machine learning, and data mining to analyze posts, comments, and interactions across social platforms. This allows businesses and researchers to identify trends, assess the effectiveness of campaigns, and predict future outcomes based on current social media activity. The information mined from these platforms offers a real-time snapshot of public opinion, which is crucial for decision-making in areas like marketing, politics, and public health.

One of the key benefits of social media mining is its ability to provide a wealth of information at a relatively low cost. Unlike traditional surveys or focus groups, social media data is generated naturally as part of users' daily activities, offering an authentic and immediate view of public attitudes. However, challenges remain in processing this data, particularly in dealing with its unstructured nature, linguistic variations, and the need for advanced tools to accurately interpret the findings.

5. Social Media Mining

Social media mining is the process of analyzing and extracting patterns from large sets of data generated by users on platforms like Facebook, Twitter, and Instagram. This form of data mining helps organizations and researchers understand the vast amount of user-generated content available online. Through techniques like sentiment analysis, social media mining can reveal public opinions, preferences, and behaviors, providing valuable insights for businesses, governments, and researchers.

Mining social media data involves using computational tools to sift through unstructured content such as posts, comments, and interactions. By applying machine learning algorithms, sentiment analysis, and network analysis, researchers can extract meaningful patterns from the data. These patterns can help identify emerging trends, track the spread of information, and gauge the public's reaction to events or products. Social media mining has become an essential tool for businesses looking to improve their customer engagement and for governments seeking to understand public sentiment on policy issues.

One of the challenges of social media mining is dealing with the sheer volume of data and the complexity of unstructured information. Users' posts often contain slang, abbreviations, and emoticons, making it difficult to accurately assess sentiment or extract reliable insights. Nevertheless, advances in NLP and machine learning continue to improve the accuracy and efficiency of social media mining, allowing organizations to better harness the potential of user-generated content for various applications.

REVIEW OF LITERATURE

Fazida Karim (2020) Social media are answerable for disturbing mental health issues. This precise investigation sums up the impacts of social system use on mental health. Fifty papers were shortlisted from google researcher databases, and after the use of different consideration and rejection models, 16 papers were picked and all papers were assessed for quality. Eight papers were cross-sectional examinations, three were longitudinal investigations, two were subjective examinations, and others were precise surveys. Discoveries were grouped into two results of mental health: tension and misery. Social media action, for example, time spent to positively affect the mental health space. Be that as it may, because of the cross-sectional structure and methodological restrictions of inspecting, there are significant contrasts. The structure of social media effects on mental health should be additionally dissected through subjective exploration and vertical accomplice contemplates.

David Thomas (2017) The reason for this examination was to more readily comprehend the connection between social media use and melancholy, by wiping out any irregularities from earlier discoveries and extending the exploration to incorporate other conceivable contributing elements that still can't seem to be investigated. Members comprised of 18-34-year-olds living in the United States. The investigation was directed through an online review on Survey Monkey. Members (N = 198) detailed that there are a few expected causal variables of sadness that outcome from the utilization of social media. These incorporate jealousy (40.45%), agitating news (15.73%), rejection (12.36%), negative posts (12.36%), clashing perspectives (8.99%), cyber bullying (3.37%), an excessive amount of time spent on social media (3.37%) and reviewing past encounters (3.37%). These outcomes affirmed that social media envy is a likely causal factor of sorrow.

PROPOSED METHODOLOGY

In proposed framework we grow new methodologies for detecting psychological disorder instances of OSN (online social network) users. We guarantee that mining social network data of people, as an inverse another to the anticipated psychological methodology conveys an exceptional chance to effectively recognize those cases at a beginning phase. In this thesis, we build up an AI system for detecting PDD (psychological disorder detection) users, in other words Social Network Mental Disorder (SNMD). This work means to construct a structure for detecting psychological disorders in social media users. We seek after to achieve our total strategy through:

1) Collection of Data
2) Cleaning and preprocessing of Data.

3) Extracting Features

Through controlling machine learning techniques with the ground truth discovered by means of the current analytical redundancy in Psychology, we remove and examine a few highlights of various classifications from OSNs, including parasocial connections, online and disconnected communication proportion, social capital, disinhibiting, self-divulgence, and bursting temporal conduct. These highlights catch significant factors or fill in as intermediaries for disorder detection.

Preprocessing Algorithms:

It is a technique that is utilized to make an interpretation of the crude data into a perfect data set. Every single time the data is gathered from various sources it is gathered in new configuration which isn't attainable for the examination. For accomplishing better outcomes from the reasonable model in Machine Learning advancements the organization of the data must be in a legitimate way. So in data preprocessing is required in light of the nearness of unformatted genuine data. In this calculation, we will talk about the a few stages associated with text preparing.

1. Stop word Removal

In this procedure stop words will be words that are generally regular in a book body and consequently considered as rather un-instructive (e.g., along these lines, and, or, the...).

One way to deal with stop word removal is to look against a language-explicit stop word reference. Another methodology is to make a stop list by arranging all words in the whole content body by consistency.

This stop list after change into a lot of non-excess words is then used to remove every one of those words from the information reports that are ordered between the top n words in this stop list.

The algorithm is executed as underneath given advances.

1) In record text is tokenized and distinct words are placed in array.
2) A single stop word is perused from stop word list.
3) The stop word is coordinated to objective content in type of array utilizing successive inquiry technique.
4) If it rises to, the word in array is removed, and the assessment is preceded with checkout length of array.
5) After removal of stop word absolutely, an extra stop word is perused from stop word list and again algorithm go to stage 2. The algorithm runs ceaselessly up to all the stop words are looked at.
6) Follow-on text void of stop words is shown, likewise required data like stop word removed, no. of stop words removed from objective content, complete include of words in target text, include of words in resultant content, separate stop word include start in objective content is shown.

2. Tokenization

In tokenization characterizes the basic process of separating a book body into discrete highlights that help as contribution for different normal language processing algorithms. Follows, tokenization processing steps:

1) Segmenting demonstration of breaking up an arrangement of arrangement into parts, for example, phrases, words, keywords, images and different highlights called tokens
2) Tokens or words are isolated by whitespace, punctuation checks or line breaks.
3) White space or punctuation imprints might be incorporated relying upon the need.
4) The tokens become the contribution for another process like parsing and text mining.

3. Stemming

Stemming is where words are decreased to a root by removing enunciation through lessening superfluous characters, normally a postfix. The outcomes can be utilized to distinguish connections and shared traits across huge datasets.

1) A stemming algorithm is a process of linguistic standardization, wherein the variation types of a word are decreased to a typical structure for example - 1. Played-play 2. Clustering–cluster
2) After stemming find in the end is that you can be improving execution of the language, while delivering an equal debasement of execution in another zone.

SYSTEM ARCHITECTURE

In a system architecture we can detect user are in pressure or not because of cooperation social network. Facebook and twitter are instances of social network. In social network individuals become all the more ready to share mind-sets on facebook. User jars various posts on a facebook. There are two kinds of data that we can use as the underlying inputs, i.e., Personal features, OSN (Online Social Network) features just as social consideration factors (being loved, remarked,) of a solitary facebook post as appeared in Fig. 1. User level posting conduct as summed up from a user month to month facebook postings, post time, post type; social connection extricated from a user's social associations with companions.

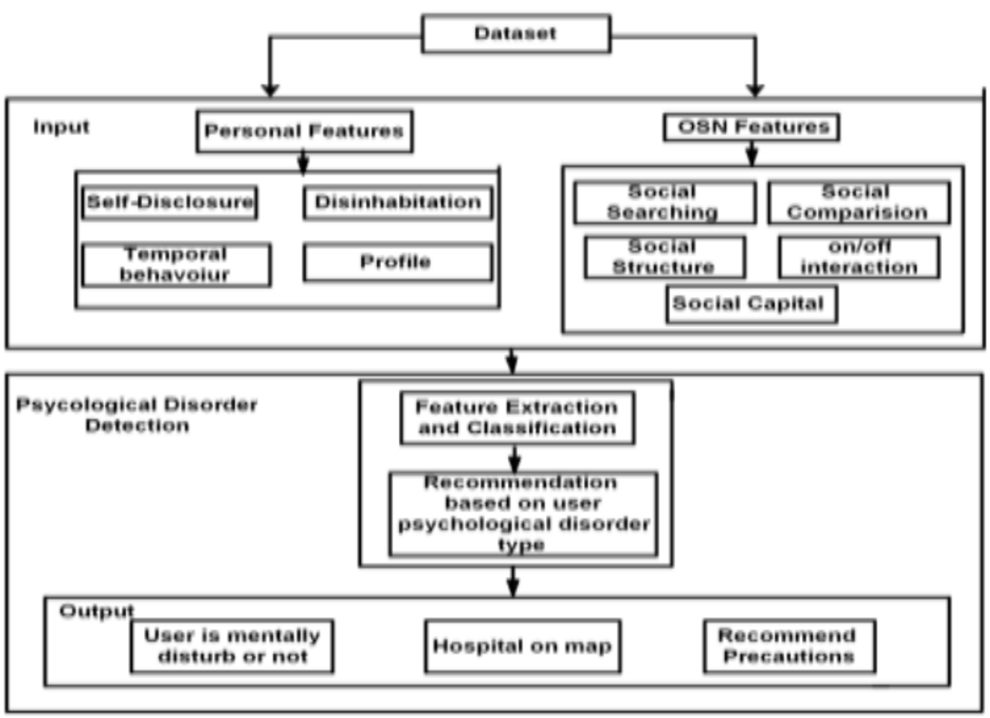

FIGURE 1: System Architecture

Our system architecture is partitioned in to the accompanying fundamental segments: collecting data, preprocessing, features extraction and grouping. We propose Psychological Disorder Detection (PDD) structure as appeared in Figure 1.

In proposed system executing an algorithm for distinguishing the psychological disorder Arrangement algorithm-naive bayes order is an overall process identified with classification, the process where items are perceived, separated and comprehended.

Algorithm : Tensor Factorization for SNMDD

Input: Tensors $\mathcal{T}$, an error threshold ϵ, and the max iteration times I_{Max}

Output: Low rank matrix $\mathbf{U}$, and matrices $\mathbf{V}$, $\mathbf{W}$, core tensor $\mathcal{C}$

1: Initialize $\mathbf{U} \in \mathbb{R}^{N\times R}$, $\mathbf{V} \in \mathbb{R}^{D\times S}$, $\mathbf{W} \in \mathbb{R}^{D\times T}$, and $\mathcal{C} \in \mathbb{R}^{R\times S\times T}$ with small random values.

2: Set η as step size

3: $d_{ii} = \sum_i z_{ij}$

4: $\mathbf{L}_z = \mathbf{D} - \mathbf{Z}$

5: **while** $t < I_{Max}$ and $Loss_t - Loss_{t+1} > \epsilon$ **do**

6:     **for** each t_{ijk} **do**

7:         Get $\nabla_{u_{i:}}\mathcal{L}$, $\nabla_{v_{j:}}\mathcal{L}$, $\nabla_{w_{k:}}\mathcal{L}$, $\nabla_{\mathcal{C}}\mathcal{L}$

8:         $\mathbf{u}_{i:}^{t+1} = \mathbf{u}_{i:}^{t} - \eta\nabla_{\mathbf{u}_{i:}^{t}}\mathcal{L}$

9:         $\mathbf{v}_{j:}^{t+1} = \mathbf{v}_{j:}^{t} - \eta\nabla_{\mathbf{v}_{j:}^{t}}\mathcal{L}$

10:        $\mathbf{w}_{k:}^{t+1} = \mathbf{w}_{k:}^{t} - \eta\nabla_{\mathbf{w}_{k:}^{t}}\mathcal{L}$

11:        $\mathcal{C}^{t+1} = \mathcal{C}^{t} - \eta\nabla_{\mathcal{C}^{t}}\mathcal{L}$

In the accompanying, we quickly sum up semi-supervised learning for SNMDD. Let xi indicate the feature vector of user i, and $x_i = u_{i:}$. Let y_i denote the class label vector of user i with length K = 3 (i.e., three types of SNMDs), and $\widehat{y}_{ik} = 1$ indicates that user i suffers from type k of SNMD. For example, {+1, −1, −1} speaks to a user with Cyber-Relationship Addiction however without Net Compulsion and Information Overload. Given a vector set D of L′ labeled samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{L}$ and L unlabeled samples $\{\mathbf{x}_j, \hat{\mathbf{y}}_j\}_{j=L+1}^{L+L'}$, the optimization issue of TSVM can be figured as follows:

$$
\begin{aligned}
\min_{\theta,\xi_{ik}} \quad & \frac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{N}\xi_{ik} + C^{*}\sum_{i=L+1}^{L+L'}\xi_{ik} \\
\text{s.t.} \quad & \mathbf{y}_{ik}(\mathbf{w}_k^T\mathbf{x}_i + b_k) \geq 1 - \xi_{ik},\ 1 \leq i \leq L, \\
& |\hat{\mathbf{y}}_{ik}(\mathbf{w}_k^T\mathbf{x}_i + b_k)| \geq 1 - \xi_{ik},\ L+1 \leq i \leq L+L', \\
& \xi_{ik} \geq 0,\ 1 \leq i \leq L+L' \text{ and } \hat{\mathbf{y}}_{ik} \in \{+1, -1\},
\end{aligned}
$$

where ξ_{ik} is the leeway variable set, and C and C^* are the tradeoff boundaries between the characterization edge and misclassification errors for marked examples and unlabeled examples, individually. The model boundaries w_k and b_k returned by this twofold learning issue speak to a double classifier related with the k-th class: $f_k(\mathbf{x}_i) = \mathbf{w}_k^T\mathbf{x}_i + b_k$. The twofold classifiers for the three SNMD types are prepared and utilized freely to anticipate the name vector $\hat{y}$ for on the other hand an unlabeled occurrence x. The test brings about the following segment find that the exactness of the semi-supervised learning without tensor is 78.3% and 83.1% for Instagram and Facebook, while STM expands the precision to 89.7% by incorporating data from Facebook and Instagram.

EXPERIMENTAL RESULTS USING TENSOR FACTORIZATION FOR SNMDD

In this area, we assess SNMDD with genuine datasets. A user concentrate with 3126 individuals is directed to assess the accuracy of SNMDD. Also, a feature study is performed. At last, we apply SNMDD for huge scope datasets and dissect the detected SNMD types.

Data Preparation and Evaluation Plan

In the following, we detail the preparation of the datasets used in our evaluation.

TABLE 1: Details of the datasets

FB_US	User profile, the friends of each user, the news feeds created by users with metadata (who likes, who comments, stickers, and geotag), the news feeds users like or comment (stickers also), events (join/decline), joined groups with events, and game posts created by game apps
IG_US	User profile, the followers/followees of each user, the media created by users with metadata (who likes, who comments, and geotag), and the contents users like or comment
FB_L	Anonymized user ID that performs the action, anonymized user ID that receives the action, and timestamp of action creation
IG_L	Anonymized media ID, anonymized ID of the user who created the media, timestamp of media creation, set of tags assigned to the media, number of likes and number of comments received

In the investigation, we initially assess the viability of proposed features, including all features (All), social collaboration features (Social) and individual profile features (Personal), with a baseline feature Duration, i.e., the all out time spent online, utilizing TSVM for semi-supervised learning in the user study. We additionally collect two enormous scope datasets, including Facebook (indicated as FB_L) with 63K hubs, 1.5M edges, and 0.84M divider posts, and Instagram (meant as IG_L) with 2K users, 9M labels, 1200M preferences, and 41M remarks. Note that some proposed features can't be separated from certain huge scope datasets, e.g., game posts and stickers are not accessible in IG_L, which is taken care of by utilizing the attribution technique. The subtleties of the data slithered from every social media are recorded in Table 3.1.

Large-Scale Experiments

To find new experiences, we apply our semi-supervised SNMDD on IG_L and FB_L to order their users and afterward dissect the detected instances of various SNMD types. Notice that the objective of this investigation is exploratory-arranged as we don't have the ground truth for the enormous datasets. We inspect whether friends of SNMD cases will in general be potential SNMD cases also. Additionally, we apply network detection on FB_L and IG_L to infer the connections between various kinds of SNMD users in their networks. At long last, the normal jump separation between the SNMD users of a similar sort is accounted for.

Evaluation of the Proposed Features

In the accompanying, we initially assess the presentation of the proposed features utilizing TSVM. We embrace Accuracy (Acc.) and Area Under Curve (AUC) for assessment of SNMDD. Additionally, Microaveraged-F1 (Micro-F1) and MacroaveragedF1 (Macro-F1) are likewise looked at for multiple-name arrangement. Table 2 sums up the normal outcomes and standard deviations, where the analyzed feature sets are signified without anyone else clarified names.

The results on the IG_US and FB_US datasets in the user study show that Duration prompts the most exceedingly terrible presentation, i.e., the results of accuracy are 34% and 36%, and the AUC are 0.362 and 0.379, individually. Utilizing (All) or parts (Social/Personal) of the features proposed in the paper beat Duration altogether. All accomplishes the best execution (78% and 83% accuracy on the IG_US and FB_US datasets, individually) on the grounds that SNMDD can catch the different features separated from data logs to adequately detect SNMD cases. Among Social and Personal, Personal beats Social in light of the fact that the features of temporal conduct (TEMP) in Personal are exceptionally viable. Since the F1 measure disregards genuine negatives, its greatness is for the most part controlled by the quantity of genuine positives, i.e., enormous classes rule little classes in microaveraging. As appeared in Table 2, Micro-F1 of Duration, Social, and Personal are bigger than Macro-F1 utilizing both IG US and FB US datasets, showing that utilizing portions of features performs better on IO and CR (enormous classes) than NC. Conversely, the exhibition of SNMDD is nearly the equivalent in Micro-F1 and Macro-F1, which demonstrates its power. The results from FB US are better than those from IG_US on the grounds that IG_US is sparser, e.g., there are no occasion and game posts on Instagram.

TABLE 2: Performance comparisons on the IG_US and FB_US datasets

	Instagram				Facebook			
Measure	Duration	Social	Personal	All	Duration	Social	Personal	All
Acc.	0.34±0.02	0.59±0.01	0.69±0.03	**0.78±0.02**	0.36±0.01	0.65±0.05	0.73±0.02	**0.83±0.02**
AUC	0.36±0.02	0.61±0.01	0.74±0.01	**0.79±0.01**	0.37±0.01	0.68±0.01	0.77±0.02	**0.84±0.01**
Micro-F1	0.42±0.02	0.71±0.01	0.78±0.04	**0.85±0.01**	0.44±0.04	0.74±0.02	0.81±0.01	**0.89±0.01**
Macro-F1	0.33±0.01	0.64±0.01	0.73±0.02	**0.85±0.01**	0.35±0.02	0.68±0.02	0.77±0.03	**0.90±0.01**

Subsequent to contrasting the results from SNMDD and the ground truth got by means of user study, we see that some bogus constructive users are detected as NC, most likely in light of the fact that individuals with NC are bound to shroud their genuine use time, e.g., the game logs of certain individuals with NC are covered up. As a result, a couple of ordinary users might be inaccurately detected as NC. Nonetheless, SNMDD by and large performs very well for NC because of some compelling features. For instance, users of NC are generally less parasocial since they are less incessant to cooperate with friends. Besides, since the NC users' friends with game advantages as a rule don't have the foggiest idea about the NC users' different friends (e.g., associates), their clustering coefficients are lower than the ordinary users.

TABLE 3: Comparisons of SNMDD with different classification techniques

Technique	Acc.	AUC
Single-source (FB)		
J48 Decision Tree Learning	74.4%	0.750
ℓ_1-regularized ℓ_2-loss SVM	77.6%	0.781
ℓ_2-regularized ℓ_2-loss SVM	77.9%	0.783
ℓ_1-regularized logistic regression	76.3%	0.776
ℓ_2-regularized logistic regression	76.4%	0.777
DTSVM	76.4%	0.774
TSVM	**83.1%**	**0.842**
Multi-source (FB+IG)		
CF	75.5%	0.759
Tucker	85.6%	0.872
STM	**89.7%**	**0.926**

Next, we contrast the proposed STM and two baseline algorithms, i.e., CF and Tucker, to incorporate features separated from both of the IG_US and FB_US datasets for learning of order models utilizing TSVM. Likewise as appeared in Table 3.3, the accuracy and AUC of STM are 89.7% and 0.926. The results show that STM, by means of the decomposed latent factor matrix U, can recoup some missing features and give extra latent data to all the more likely portray the users. Interestingly, CF, which basically links features from FB_US and IG_US, plays out the most noticeably awful. Its accuracy and AUC are 75.5% and 0.759, separately, surprisingly more dreadful than those of some single-source learning algorithms utilizing just the FB_US dataset, e.g., 77.9% and 0.783 for l_2-regularized l_2-misfortune SVM. This is on the grounds that CF loses the relationships in certain features and along these lines presents commotions. Then again, while Tucker can accomplish 85.6% accuracy and 0.872 AUC, its presentation is still not as extraordinary as STM. This result shows that STM, by consolidating significant earlier information about qualities of SNMD cases (see conversation with respect to Eq. (3.2)), can determine more exact and precise latent features than Tucker to accomplish the best execution in SNMDD.

CONCLUSION

This study presents a comprehensive system for detecting mental health disorders using data mined from social media platforms. By leveraging machine learning techniques, particularly Support Vector Machines, the system effectively identifies users exhibiting psychological disorders based on their online behaviors. The integration of social interaction features and personal profile data significantly enhances detection accuracy. Results from large-scale datasets demonstrate the system's potential for real-world applications in monitoring mental health through digital platforms. Future research should explore refining feature extraction methods and expanding the system to include more social media platforms for broader applicability.

REFERENCES

- Chi, Yun, Xiaodan Song, Dengyong Zhou, Koji Hino, and Belle L. Tseng. 2007. Evolutionary Spectral Slustering by Incorporating Temporal Smoothness. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. 153-162.

- Chisari, M. (2009). The future of social networking. In Proceedings of the W3C workshop on the future of social networking
- Cyworld, About Cyworld(2007) at 14 January 2008.
- Du, W., Deng, J., Han, Y., Chen, S. and Varshney, P. (2004) 'A key management scheme for wireless sensor networks using deployment knowledge', INFOCOM 2004, Vol. 1, IEEE, pp.586–597.
- Dutton, W. H. (2004). Social transformation in an information society: rethinking access to you and theworld. Retrieved from UNESCO Archives English - UNESCO HQ Social Sciences SSDCN (stock2E)
- Erétéo, G., Buffa, M., Gandon, F., Leitzelman, M., &Limpens, F. (2009). Leveraging social data with semantics. In Proceedings of the W3C workshop on the future of social networking
- Gross, R., &Acquisti, A. (2005). Information Revelation and Privacy in Online Social Networks (The Facebook case). ACM Workshop on Privacy in the Electronic Society (WPES), 2005. Retrieved from http://www.heinz.cmu.edu/~acquisti/papers/privacyfacebook-gross-acquisti.pdf
- H. Gao, J. Hu, C.Wilson, Z. Li, Y. Chen, and B. Zhao, "Detecting and Characterizing Social Spam Campaigns," in Internet Measurement Conference (IMC), 2010.
- H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards Online Spam Filtering in Social Networks," in Symposium on Network and Distributed System Security (NDSS), 2012
- Ho, A., Maiga, A., &Aïmeur, E. (2009, May). Privacy protection issues in social networking sites. In Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference
- Irani, D., Web, S., et al.: Large online social footprints-an emerging threat. In: 2009 IEEE International Conference on Computational Science and Engineering, pp. 271–277 (2009) 15(3), 28-34.
- J. Baltazar, J. Costoya, and R. Flores, "KOOBFACE: The Largest Web 2.0 Botnet Explained," 2009.
- J. Bonneau, J. Anderson, R. Anderson, and F. Stajano, "Eight friends are enough: social graph approximation via public listings," in Proceedings of the Second ACM EuroSys Workshop on Social Network Systems. ACM, 2009, pp. 13–18.
- . J. M. Kleinberg. Challenges in mining social network data: processes, privacy, and paradoxes. In KDD '07:Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining,pages 4–5, New York, NY, USA, 2007. ACM.
- Joshi, Y., Saklikar, S.: PhishGuard: A browser plug-in for protection from phishing. In: IMSAA (2008)
- Kim, Min-Soo, and Jiawei Han. 2009. A Particle-and-Density Based Evolutionary Clustering Method for Dynamic Networks. Proceedings of the VLDB Endowment, 2(1): 622-633.
- Kimura, Masahiro, and Kazumi Saito. 2006. Tractable Models for Information Diffusion in Social Networks. In Knowledge Discovery in Databases: PKDD 2006, Springer Berlin Heidelberg, 259-271.

- Kunegis, J., Lommatzsch, A., &Bauckhage, C. (2009, April). The slashdot zoo: mining a social network with negative edges. In Proceedings of the 18th international conference on world wide web
- Leskovec, J., Huttenlocher, D. P. & Kleinberg, J. M, Governance in social media: A case study of the wikipedia promotion process In ICWSM, 2010 April.
- Leskovec, Jure, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-Effective Outbreak Detection in Networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 420-429.
- Schwartz HA, Eichstaedt J, Kern M, Park G, Sap M, Stillwell D, Kosinski M, Ungar L, "Towards assessing changes in degree of depression through facebook", Proceedings of the Workshop on Computational Linguistics and Clinical Psychology, From Linguistic Signal to Clinical Reality ACL, 2014, pp.118-125.
- Moreno M. A and Kolb J," Social networking sites and adolescent health. Pediatric Clinics of North America", 2012.
- Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ," Forecasting the Onset and Course of Mental Illness with Twitter Data.", 2016 .