Detection of Social Network Mental Disorders Through Mining of Online Social Media

Sindhuja Enugula[1], Shireen Begum[2], Deepak Kumar Gandham[3], K. VenkataRamu[4]

[1,2,3]UG Student, [4]Assistant Professor, Department of CSE
[1,2,3,4]Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Hyderabad, India.

ABSTRACT

The explosive growth in popularity of social networking leads to the problematic usage. An increasing number of social network mental disorders (SNMDs), such as Cyber-Relationship Addiction, Information Overload, and Net Compulsion, have been recently noted. Symptoms of these mental disorders are usually observed passively today, resulting in delayed clinical intervention. In this paper, we argue that mining online social behavior provides an opportunity to actively identify SNMDs at an early stage. Itis challenging to detect SNMDs because the mental status cannot be directly observed from online social activity logs. Our approach, new and innovative to the practice of SNMD detection, does not rely on self-revealing of those mental factors via questionnaires in Psychology. Instead, we propose a machine learning framework, namely, logistic regression model (LRM), that exploits features extracted from social network data to accurately identify potential cases of SNMDs. We also exploit the comparative evaluation with K-nearest neighbour (KNN) classifier.

Keywords: machine learning, social network mental disorder, KNN classifier, logistic regression model.

1. INTRODUCTION

With the explosive growth in popularity of social networking and messaging apps, online social networks (OSNs) have become a part of many people's daily lives. Most research on social network mining focuses on discovering the knowledge behind the data for improving people's life. While OSNs seemingly expand their users' capability in increasing social contacts, they may decrease the face-to-face interpersonal interactions in the real world. Due to the epidemic scale of these phenomena, new terms such as Phubbing (Phone Snubbing) and Nomophobia (No Mobile Phone Phobia) have been created to describe those who cannot stop using mobile social networking apps. In fact, some social network mental disorders (SNMDs), such as Information Overload and Net Compulsion [1], have been recently noted. For example, studies point out that 1 in 8 Americans suffer from problematic Internet use. Moreover, leading journals in mental health, such as the American Journal of Psychiatry [2], have reported that the SNMDs may incur excessive use, depression, social withdrawal, and a range of other negative repercussions. Indeed, these symptoms are important components of diagnostic criteria for SNMDs [3] e.g., excessive use of social networking apps – usually associated with a loss of the sense of time or a neglect of basic drives, and withdrawal – including feelings of anger, tension, and/or depression when the computer/apps are inaccessible. SNMDs are social-oriented and tend to happen to users who usually interact with others via online social media. Those with SNMDs usually lack offline interactions, and as a result seek cyber-relationships to compensate. Today, identification of potential mental disorders often falls on the shoulders of supervisors (such as teachers or parents) passively. However, since there are very few notable physical risk factors, the patients usually do not actively seek medical or psychological services. Therefore, patients would only seek clinical interventions when their conditions become very severe. However, a recent study shows a strong correlation between suicidal attempt and SNMDs [4], which indicates that adolescents suffering from social network addictions have a much higher risk of suicidal inclination than non-addictive users. The research also reveals that social network addiction may negatively impact emotional status, causing higher hostility, depressive mood, and compulsive behavior. Even more alarming is that the delay of early intervention may seriously damage individuals' social functioning. In short, it is desirable to have the ability to actively detect potential SNMD users on OSNs at an early stage. Although previous work in Psychology has identified several crucial mental factors related to SNMDs, they are mostly examined as standard diagnostic criteria in survey questionnaires. To automatically detect potential SNMD cases of OSN users, extracting these factors to assess users' online mental states is incredibly challenging. For example, the extent of loneliness and the effect of disinhibition of OSN users are not easily observable.3 Therefore, there is a need to develop new approaches for detecting SNMD cases of OSN users. We argue that mining the social network data of individuals as a complementary alternative to the conventional psychological approaches provides an excellent opportunity to actively identify those cases at an early stage. In this paper, we develop a machine learning framework for detecting SNMDs, which we call LRM based Social Network Mental Disorder Detection (SNMDD).

Specifically, we formulate the task as a semi-supervised classification problem to detect three types of SNMDs [1]:

i) Cyber-Relationship Addiction, which shows addictive behavior for building online relationships.

ii) Net Compulsion, which shows compulsive behavior for online social gaming or gambling.

iii) Information Overload, which is related to uncontrollable surfing.

By exploiting machine learning techniques with the ground truth obtained via the current diagnostic practice in Psychology [1], we extract and analyse the following crucial categories of features from OSNs: 1) social comparison, 2) social structure, 3) social diversity, 4) parasocial relationships, 5) online and offline interaction ratio, 6) social capital, 7) disinhibition, 8) self-disclosure, and 9) bursting temporal behavior. These features capture important factors or serve as proxies for SNMD detection. For example, studies manifest that users exposed to positive posts from others on Facebook with similar background are inclined to feel malicious envy and depressed due to the social comparison. The depression leads users to disorder behaviours, such as information overload or net compulsion. Therefore, we first identify positive newsfeeds and then calculate the profile similarity and relation familiarity between friends. As another example, a parasocial relationship is an asymmetric interpersonal relationship, i.e., one party cares more about the other, but the other does not. This asymmetric relationship is related to loneliness, one of the primary mental factors pushing users with SNMDs to excessively access online social media [5]. Therefore, we extract the ratio of the number of actions to and from friends of a user as a feature. In this paper, the extracted features are carefully examined through a user study.

2. RELATED WORK

Internet Addiction Disorder (IAD) is a type of behavior addiction with the patients addicted to the Internet, just like those addicting to drugs or alcohol [3]. Many research works in Psychology and Psychiatry have studied the important factors, possible consequences, and correlations of IAD [10]. King et al.[11] investigate the problem of simulated gambling via digital and social media to analyse the correlation of different factors, e.g., grade, ethnicity. Li et al. [12] examine the risk factors related to Internet addiction. Kim et al. [13] investigate the association of sleep quality and suicide attempt of Internet addicts. On the other hand, recent research in Psychology and Sociology reports several mental factors related to social network mental disorders. Research indicates that young people with narcissistic tendencies and shyness are particularly vulnerable to addiction with OSNs. However, the above research explores various negative impacts and discusses potential reasons for Internet addiction. By contrast, this paper proposes to automatically identify SNMD patients at the early stage according to their OSN data with a novel tensor model that efficiently integrate heterogeneous data from different OSNs.

Research on mental disorders in online social networks receives increasing attention recently [14-16]. Among them, content-based textual features are extracted from user generated information (such as blog, social media) for sentiment analysis and topic detection. Chang et. al [14] employ an NLP-based approach to collect and extract linguistic and content-based features from online social media to identify Borderline Personality Disorder and Bipolar Disorder patients. Saha et al. [15] extract the topical and linguistic features from online social media for depression patients to analyse their patterns. Choudhury et al. [16] analyse emotion and linguistic styles of social media data for Major Depressive Disorder (MDD). However, most previous research focuses on individual behaviours and their generated textual contents but do not carefully examine the structure of social networks and potential Psychological features. Moreover, the developed schemes are not designed to handle the sparse data from multiple OSNs. In contrast, we propose a new multi-source machine learning approach, i.e., STM, to extract proxy features in Psychology for different diseases that require careful examination of the OSN topologies, such as Cyber-Relationship Addiction and Net Compulsion.

Our framework is built upon LRM, which has been widely used to analyse OSNs in many areas. In addition, we present a new tensor model that not only incorporates the domain knowledge but also well estimates the missing data and avoids noise to properly handle multi-source data. Caballero et al. [8] estimate the probability of mortality in ICU by modelling the probability of mortality as a latent state evolving over time. Zhao et al. [9] propose a hierarchical learning method for event detection and forecasting by first extracting the features from different data sources and then learning via geographical multi-level model. However, the SNMD data from different OSNs may be incomplete due to the heterogeneity. For example, the profiles of users may be empty due to the privacy issue, different functions on different OSNs (e.g., game, check-in, event),etc. We propose a novel tensor-based approach to address the issues of using heterogeneous data and incorporate domain knowledge in SNMD detection.

3. PROPOSED METHODOLOGY

In this paper, we aim to explore data mining techniques to detect three types of SNMDs [1]:

1) Cyber-Relationship (CR) Addiction, which includes the addiction to social networking, checking, and messaging to the point where social relationships to virtual and online friends become more important than real-life ones with friends and families.

2) Net Compulsion (NC), which includes compulsive online social gaming or gambling, often resulting in financial and job-related problems.

3) Information Overload (IO),which includes addictive surfing of user status and news feeds, leading to lower work productivity and fewer social interactions with families and friends offline.

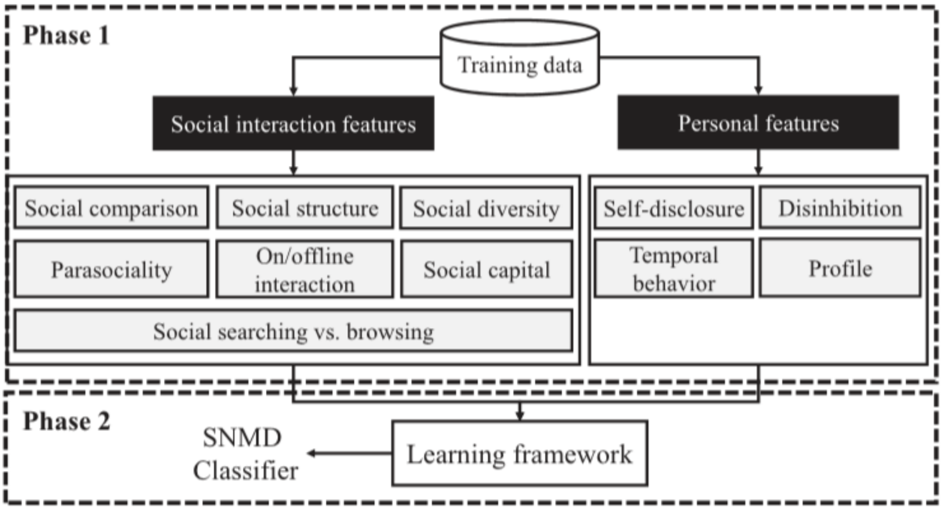

Fig. 1: Framework of SNMDD.

Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a KNN classifier and LRM. In this study, we propose a two-phase framework, called SNMDD. The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon KNN and LRM. Two key challenges exist in design of SNMDD:

- We are not able to directly extract mental factors like what have been done via questionnaires in Psychology and thus need new features for learning the classification models.
- We aim to exploit user data logs from multiple OSNs and thus need new techniques for integrating multi-source data based on SNMD characteristics.

3.1. K-Nearest Neighbour

K-Nearest neighbour is a lazy learner technique. This algorithm depends on learning by analogy. It is a supervised classification method. This classifier is used extensively for classification purpose. This classifier waits till the last minute prior to build some model on a specified tuple as compared to earlier classifiers. The training tuples are characterized in N-dimensional space in this classifier. This classification model looks for the k training tuples nearest to the indefinite sample in case of an indefinite tuple. Then, this classifier puts the sample in the closest class.

Disadvantages

Results with less accuracy as low as 50% due to following:

- **Does not work well with large dataset:** In large datasets, the cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithm.
- **Does not work well with high dimensions:** The KNN algorithm does not work well with high dimensional data because with large number of dimensions, it becomes difficult for the algorithm to calculate the distance in each dimension.
- **Need feature scaling:** We need to do feature scaling (standardization and normalization) before applying KNN algorithm to any dataset. If we do not do so, KNN may generate wrong predictions.
- **Sensitive to noisy data, missing values, and outliers:** KNN is sensitive to noise in the dataset. We need to manually impute missing values and remove outliers.

3.2. Logistic Regression

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$\frac{1}{(1 + e^{-value})}$$

Where e is the base of the natural logarithms (Euler's number or the exp () function in your spreadsheet) and value is the actual numerical value that you want to transform.

Advantages

- It performs well when the dataset is linearly separable.
- This is less prone to over-fitting, but it can overfit in high dimensional datasets. You should consider Regularization (L1 and L2) techniques to avoid over-fitting in these scenarios.
- It is not only giving a measure of how relevant a predictor (coefficient size) is, but also its direction of association (positive or negative).
- This is easier to implement, interpret and very efficient to train.

ISSN: 2278-4632
Vol-10 Issue-6 No. 1 June 2020

4. RESULTS AND DISCUSSION

In this section, we evaluate SNMDD with real datasets. A user study with 3126 people is conducted to evaluate the accuracy of SNMDD. Moreover, a feature study is performed. Finally, we apply SNMDD on large-scale datasets and analyse the detected SNMD types.

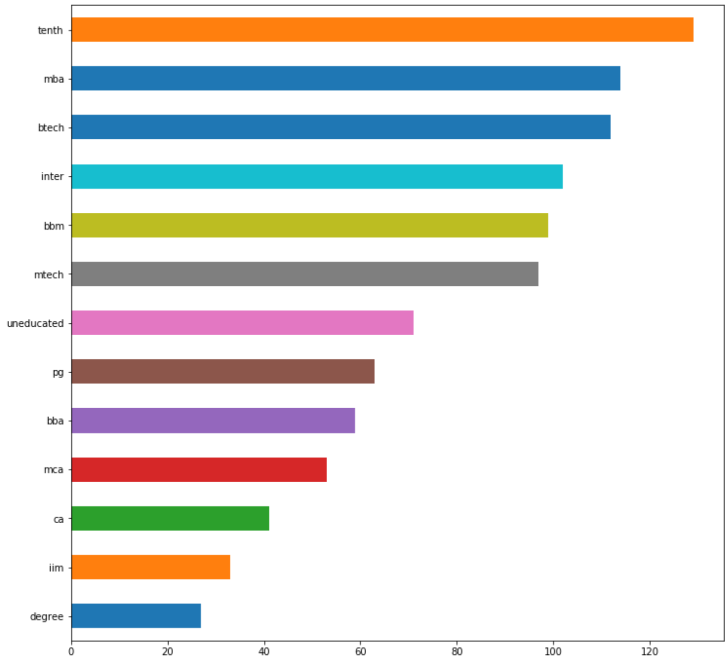

Fig. 2: Dataset of students from different courses.

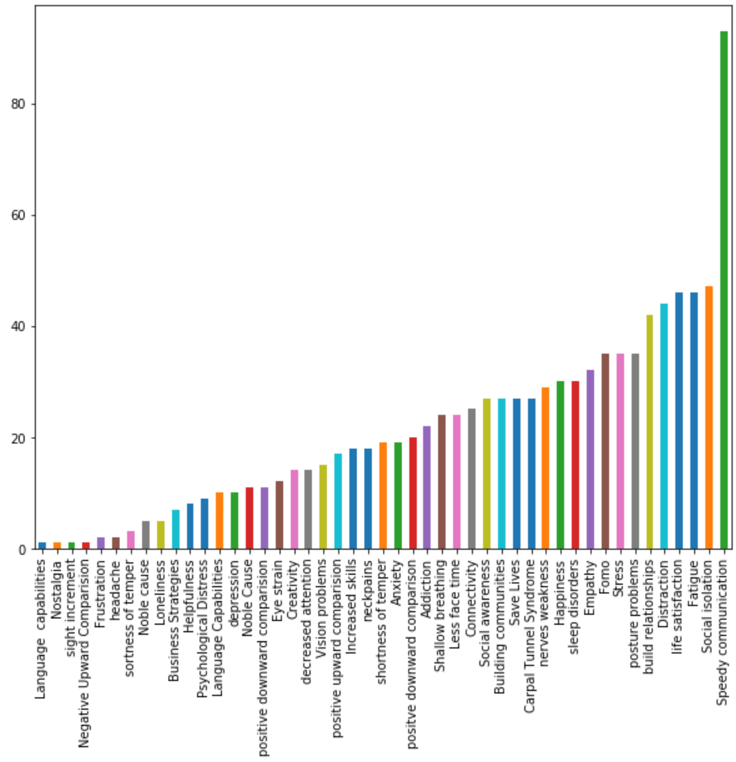

Fig. 3: Classification of disorders based on OSNs.

ISSN: 2278-4632
Vol-10 Issue-6 No. 1 June 2020

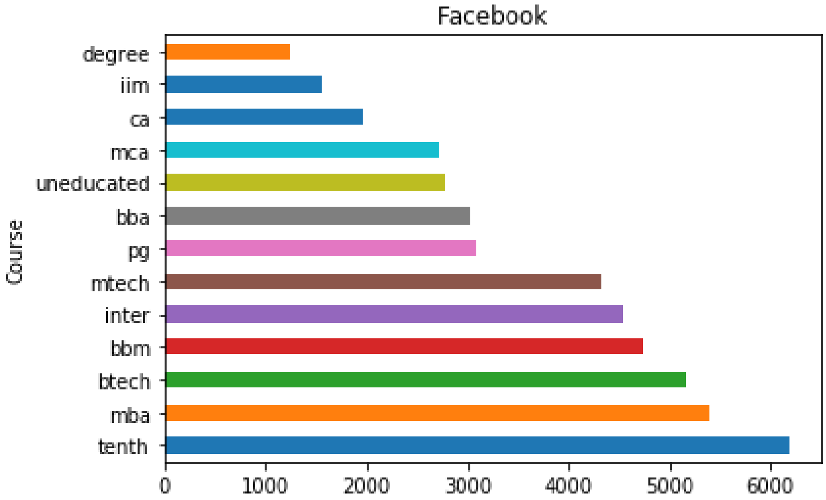

Fig. 4: Data from the Facebook.

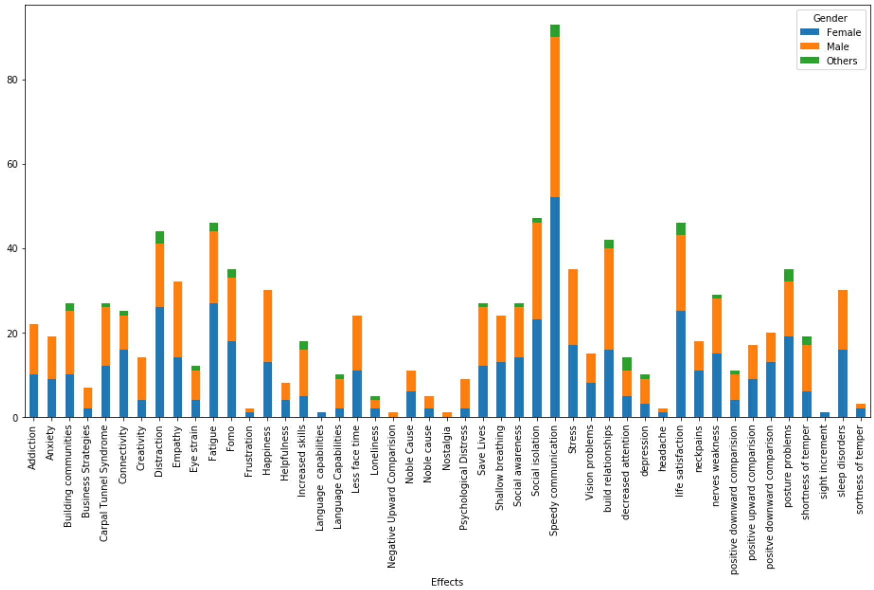

Fig. 5: Type of disorder with gender classification.

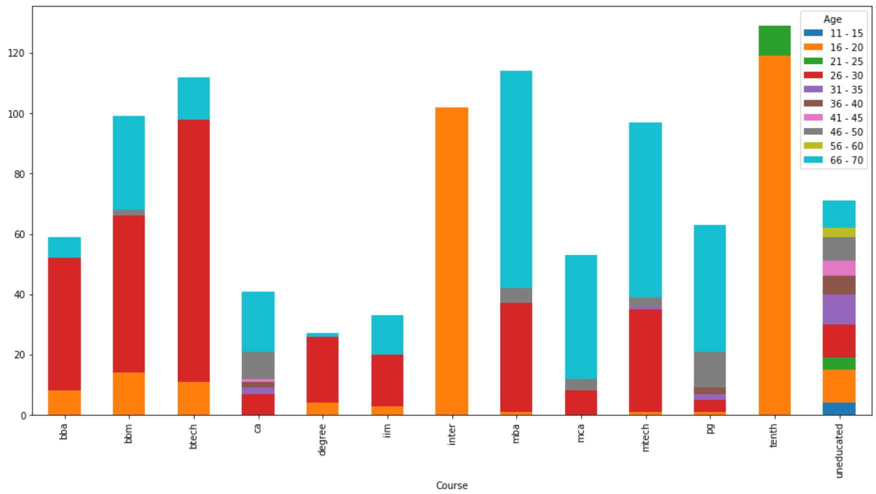

Fig. 6: Classification of SNMD with age.

ISSN: 2278-4632
Vol-10 Issue-6 No. 1 June 2020

5. CONCLUSIONS

In this paper, we try to automatically identify potential online users with SNMDs. We propose an SNMDD framework that explores various features from data logs of OSNs and a new tensor technique for deriving latent features from multiple OSNs for SNMD detection. This work represents a collaborative effort between computer scientists and mental healthcare researchers to address emerging issues in SNMDs. As for the next step, we plan to study the features extracted from multimedia contents by techniques on NLP and computer vision. We also plan to further explore new issues from the perspective of a social network service provider, e.g., Facebook or Instagram, to improve the well-beings of OSN users without compromising the user engagement.

REFERENCES

[1] K. Young, M.Pistner, J.O'Mara, and J. Buchanan. Cyber-disorders: The mental health concern for the new millennium. Cyberpsychol. Behav., 1999.
[2] J. Block. Issues of DSM-V: internet addiction. American Journal of Psychiatry, 2008.
[3] K. Young. Internet addiction: the emergence of a new clinical disorder, Cyberpsychol. Behav., 1998.
[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., 2014.
[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., 2013.
[6] D. La Barbera, F. La Paglia, and R. Valsavoia. Social network and addiction. Cyberpsychol. Behav., 2009.
[7] K. Chak and L. Leung. Shyness and locus of control as predictors of internet addiction and internet use. Cyberpsychol. Behav., 2004.
[8] K. Caballero and R. Akella. Dynamically modeling patients health state from electronic medical records: a time series approach. KDD, 2016.
[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, 2016.
[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, 2013.
[11] D. L. King, P. H. Delfabbro, D. Kaptsis, and T. Zwaans. Adolescent simulated gambling via digital and social media: an emerging problem. Computers in Human Behavior, 2014.
[12] D. Li, X. Li, L. Zhao, Y. Zhou, W. Sun, and Y. Wang. Linking multiple risk exposure profiles with adolescent Internet addiction: insights from the person-centered approach. Computers in Human Behavior, 2017.
[13] K. Kim, H. Lee,J. P. Hong, M.J. Cho, M. Fava,D. Mischoulon, D. J. Kim, and H. J. Jeon. Poor sleep quality and suicide attempt among adults with internet addiction: a nationwide community sample of Korea. PLOS ONE, 2017.

[14] C.-H Chang, E. Saravia, and Y.-S. Chen. Subsonscious crowdsourcing: a feasible data collection mechanismfor mental disorder detection on social media, ASONAM, 2016.
[15] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh. A framework for classifying online mental health-related communities with an interest in depression. IEEE Journal of Biomedical and Health Informatics, 2016.
[16] M. Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. ICWSM, 2013.