Mental Health Prediction Using Data Mining

I. G. Hemanandhini and C. Padmavathy

Abstract Mental illness is a condition that affects the behaviour, attitude and mannerisms of a person. They are highly common in these days of isolation due to the on-going pandemic. Almost 450 million people worldwide suffer from some kind of mental illness. Mental health problems do not only affect adults, but also it has significant impact on kids and teenagers. It is totally normal and understandable to experience fear during the time of COVID-19 pandemic. Loneliness, isolation, unhealthy alcohol and substance usage, self-harm or suicidal behaviour are all projected to escalate as new policies and impacts are implemented, especially quarantine and its effects on many people's usual habits, schedules or livelihoods. Furthermore, psychiatric disorders have become one of the most severe and widespread public health issues. Early diagnosis of mental health issues is critical for further understanding mental health disorders and providing better medical care. Unlike the diagnosis of most health diseases, which is dependent on laboratory testing and measures, psychiatric disorders are usually classified based on a person's self-report of detailed questionnaires intended to identify specific patterns. The project would use a person's tweets, a few customized questions and answers, and a few personal data to measure a person's mental well-being ranking. This initiative would be immensely helpful to anyone who uses social media sites on a regular basis in order to live a stress-free life and diagnose mental health problems before they get too serious.

Keywords Mental illness · Psychiatric disorders · Mental well-being · COVID-19

I. G. Hemanandhini (✉) · C. Padmavathy
Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India
e-mail: hemanandhini@srec.ac.in

C. Padmavathy
e-mail: padma.dhanush@srec.ac.in

S. Smys et al. (eds.), *Inventive Computation and Information Technologies*, Lecture Notes in Networks and Systems 336, https://doi.org/10.1007/978-981-16-6723-7_52

1 Introduction

COVID-19 has resulted in quarantining or isolating as a new trend that has a very passive effect on individuals. Individuals may develop thoughts such as depression, anxiety and suicidal ideation as a result of this. Workplaces and educational institutions are doing their best to address the situation, but it is insufficient. The challenge is to develop a model that can accurately predict an individual's mental health state and assist in monitoring and curing it at an earlier stage.

The primary goal of this paper is to anticipate a person's mental well-being using social media platforms, specifically Twitter. Based on the individual's tweets, this paper will predict a mental health score. It will be extremely beneficial to those who use social media platforms on a daily basis and will assist them in monitoring their mental health in order to live a stress-free life.

2 Related Work

This section briefs about the related works carried out for predicting a person's mental well-being via social platforms.

S. E. Jordan et al. conducted a survey dictating the use of Twitter data for predicting public health. Here, various methods were used for mining the Twitter data for public health monitoring. Research papers where Twitter data is classified as users or tweets are considered for the survey for monitoring the health of persons in a better way. Also, papers published from 2010 to 2017 were taken for conducting survey. The approaches used to categorize the Twitter content in many ways are distinguished. While it is difficult to compare research, since there are so many various ways for using Twitter and interpreting data, this state-of-the-art review highlights the huge potential of using Twitter for public health surveillance.

Heiervang et al. conducted a structured psychiatric interview for the parents for predicting the child mental state. Parents were interviewed face to face in 2003, and they finished the interview online in 2006. Interviews were preceded by printed questionnaires covering child and family variables in both surveys. Web-based surveys can be completed more quickly and at a lesser cost than traditional methods including personal interviews. Point estimates of psychopathology appear to be particularly vulnerable to selective participation although patterns of connections appear to be more durable.

3 Proposed System

Coronavirus has crossed the globe, isolating or disengaging numerous people, bringing about antagonistic psychological well-being impacts for some like uneasiness, melancholy, self-destruction and self-hurt. Working environments/educational establishments that encourage mental prosperity and help individuals with mental incapacities are bound to limit non-attendance, improve profitability and receive the expert and individual rewards that accompany it. The test is to make a model that will foresee the emotional wellness of people and accordingly help psychological well-being suppliers to convey forward with the treatment in this period of scarcity.

These days, most of the mental health problems are identified and treated at later stage. We propose a unique technique to mental health detection using user tweets as an early discovery system to actively identify probable mental health situations. A machine learning framework has been developed for finding a person's mental well-being. We analyse a person's tweets, along with a few of their personal details and predict whether or not the person should see a therapist based on a series of quizzes. The proposed approach employs naïve Bayes and linear regression techniques to find a person's mental health by means of their tweets. To improve efficiency, we perform a quiz analysis with a decision tree algorithm and predict the scores.

4 Algorithm Description

4.1 Naïve Bayes Algorithm

It is a method based on Bayes' theorem and the assumption that indicators are autonomous. A naïve Bayes classifier, in simple terms, assumes that the presence of one variable in a class has no influence on the presence of another. For example, if a natural product is red, oval and around 3 creeps across, it is considered an apple. Regardless of whether these characteristics rely on one another or on the existence of other characteristics, they all contribute to the probability that this natural commodity is an apple, which is why it is regarded as 'Credulous'. The Bayes model is simple to construct and is particularly useful for extremely large informative indexes. Along with its simplicity, naïve Bayes is considered to outperform even the most sophisticated order techniques. From $P(c)$, $P(x)$ and $P(x|c)$, the Bayes hypothesis provides a method for determining back probability $P(c|x)$.

4.2 Linear Regression Algorithm

Linear regression is an AI calculation that is based on learned data. It performs a relapse simulation. A regression model is used to model an objective expectation

esteem that is contingent on free variables. It is primarily used for evaluating and exploring the relationship between variables. Different relapse models differ in terms of the type of relationship; they consider between reliant and autonomous factors as well as the number of free factors they employ.

Linear regression enacts the task of predicting the value of a dependent variable (y) in light of a given autonomous variable (x). In this way, this relapse protocol discovers a direct link between x (input) and y (output). Linear regression is the name given to it as a result.

4.3 Decision Tree

The decision tree algorithm is part of the supervised learning algorithms family. The decision tree algorithm, unlike other supervised learning algorithms, can also be used to solve regression and classification problems.

The global variables we have determines the different sorts of decision trees we have. There are two forms of it:

Categorical Variable Decision Tree: A categorical variable decision tree is a decision tree with a categorical target variable.

Continuous Variable Decision Tree: A continuous variable decision tree is one that has a continuous focus variable (Fig. 1).

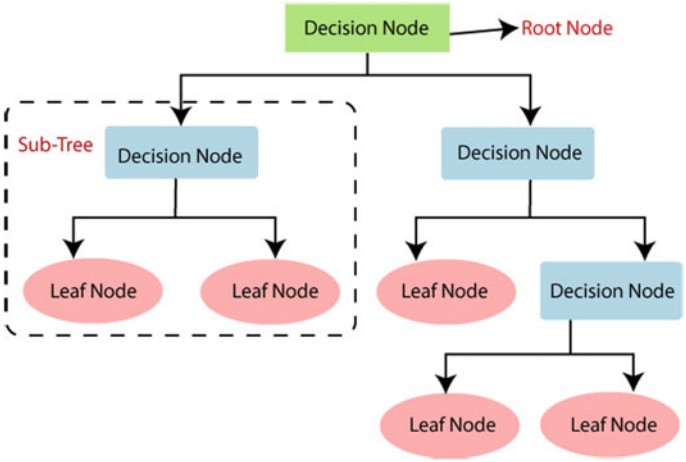

Fig. 1 Decision tree

5 Implementation

Today, psychological wellness is anticipated at a later stage. To effectively distinguish expected psychological well-being by mining information logs of online media clients as an early discovery framework, we present a novel way for recognizing emotional well-being. We foster an AI structure to distinguish emotional wellness. The proposed approach can be communicated to provide early notification of anticipated patients. We analyse the client's tweets and apply naïve Bayes and linear regression calculations to get the clients OCEAN examination, i.e. openness, conscientiousness, extraversion, agreeableness and neuroticism. We utilize a test investigation to become familiar with the client and a choice tree calculation to figure their emotional wellness score. Alongside these two qualities, we get some close to home data, and the emotional wellness score is anticipated, alongside a message showing whether the individual should see a therapist.

5.1 Module Description

1. **OCEAN Analysis**: In this module, we take the tweets of the users as the datasets and process the datasets to get the OCEAN (openness, conscientiousness, extraversion, agreeableness, neuroticism) analysis of the user. The tweets are hence cleaned and algorithms such as naïve Bayes and linear regression are performed to calculate the emotion of the tweets.
2. **Quiz Analysis**: The proposed quiz analysis consists of 20 different customized questions that will help us to give a clearer analysis of the mental state. We apply a decision tree algorithm to the answers and predict the score.
3. **Prediction of Mental Health**: Here, we predict the final score using both the scores obtained from ocean analysis and quiz analysis. We suggest depending on the score whether a person needs to consult the therapist or not.

5.2 System Architecture

See Fig. 2.

6 Result and Discussions

In this paper, we used different algorithms to try to determine the mental health of people who use social media sites, mostly Twitter. We successfully predicted the user's mental health score and recommended whether or not the individual should see a therapist. Using naïve Bayes, linear regression and decision tree on their tweets, we

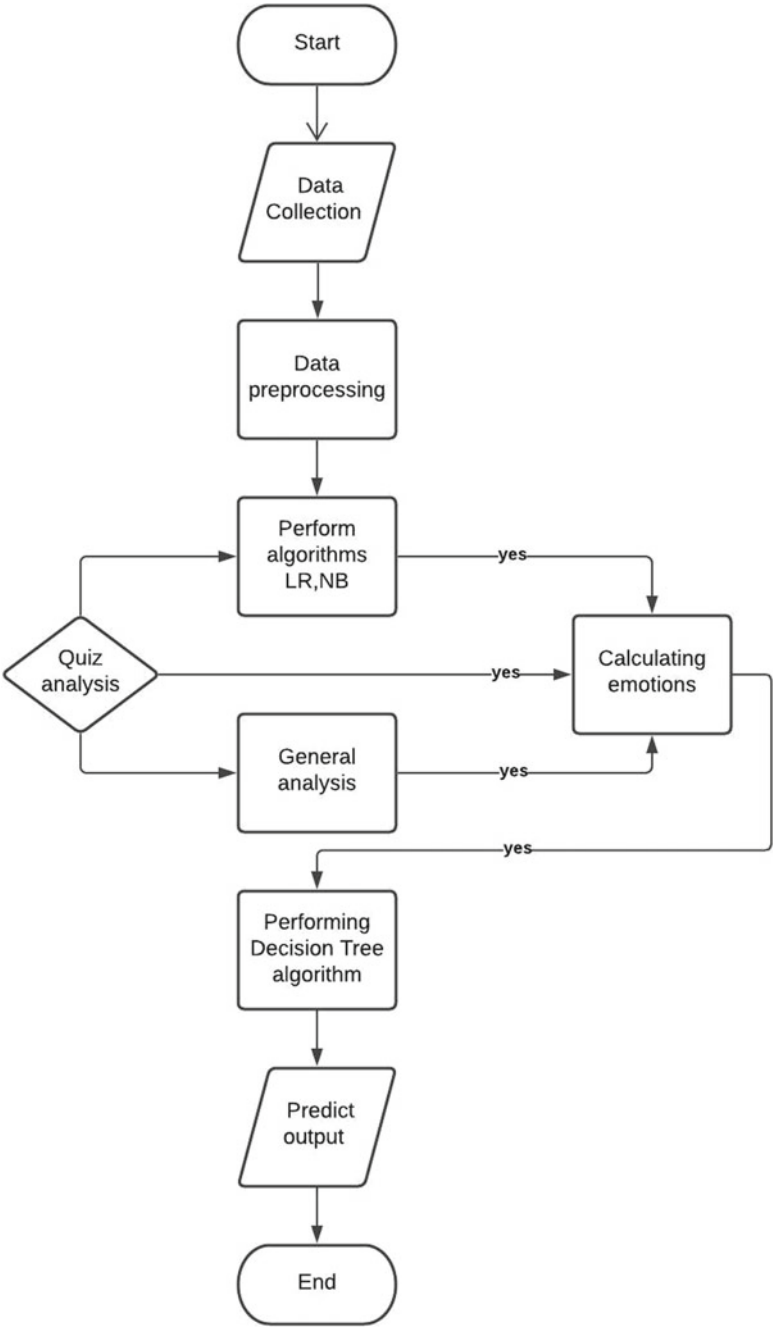

Fig. 2 System architecture

performed OCEAN analysis and got an output with greater accuracy. Another such output was predicted using a survey with several questions based on the candidate's behaviour. Personal details like work scenarios and family history were taken into account as well. These outputs were then added and taken average of. The following result is more accurate than the previous works based on various parameters (Figs. 3, 4, 5, 6, 7, and 8).

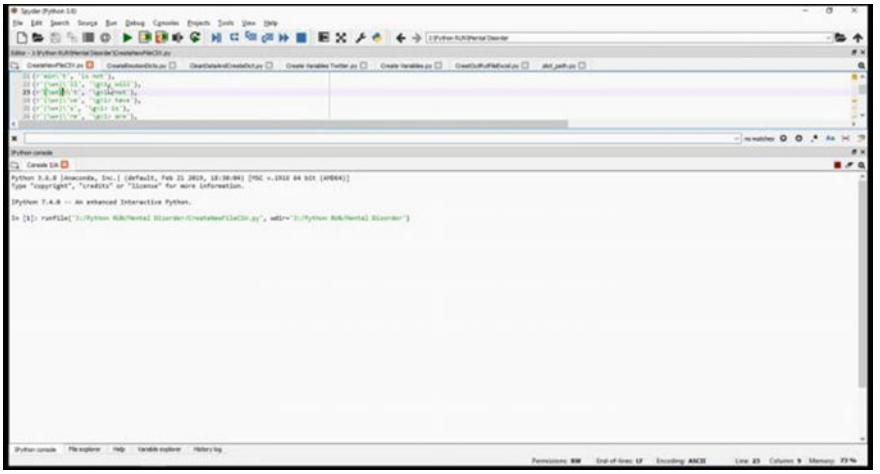

Fig. 3 Data pre-processing

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Post	Months	Hours	anger	anticipatio	disgust	fear	joy	openness	conscientio	extraversio	agreeable	neuroticisr
2	b7b7764cf	likes soun	3	15	0	0	0	0	0	0	0	0	0	0
3	b7b7764cf	sleepy not	3	08	0	0	0	0	0	0	0	0	0	0
4	b7b7764cf	sore wants	3	13	2	0	0	1	0	2	0	2	0	1
5	b7b7764cf	likes day s	3	04	0	0	0	0	0	0	0	0	0	0
6	b7b7764cf	home 3	3	02	0	0	0	0	0	0	0	0	0	0
7	b7b7764cf	wwwthejok	3	15	0	0	0	0	0	0	0	0	0	0
8	b7b7764cf	saw nun z	3	05	0	0	0	0	0	1	0	0	0	1
9	b7b7764cf	kentucky 4	3	06	0	1	0	1	1	0	1	0	0	0
10	b7b7764cf	finish digit	3	14	1	1	0	1	0	2	1	0	1	1
11	b7b7764cf	celebrating	3	23	0	1	0	0	2	0	2	1	0	0
12	b7b7764cf	crush gree	3	19	0	0	0	0	1	0	1	0	0	1
13	b7b7764cf	magic brai	3	04	0	0	0	0	0	0	0	0	0	0
14	b7b7764cf	saw transf	3	04	0	1	0	0	1	1	1	1	1	1
15	b7b7764cf	wants mee	3	03	0	0	0	0	0	0	0	0	0	0
16	b7b7764cf	desires thr	3	21	1	3	0	1	3	0	3	0	2	1
17	b7b7764cf	going bed	3	01	0	0	0	0	0	0	0	0	0	0
18	b7b7764cf	reading ad	3	22	0	0	0	0	0	0	1	0	0	0
19	b7b7764cf	thinks inta	3	02	0	0	0	0	0	0	1	0	0	0
20	b7b7764cf	tired let go	3	05	0	0	0	0	0	1	0	0	0	0
21	b7b7764cf	discovering	3	06	0	0	0	0	0	0	0	0	0	0
22	b7b7764cf	watching c	3	04	0	1	0	1	0	0	0	0	0	0
23	b7b7764cf	getting urg	3	00	0	0	0	0	0	0	0	0	0	1
24	b7b7764cf	woulda tho	3	03	0	1	0	1	0	0	0	0	0	0
25	b7b7764cf	wishes dev	4	02	0	1	0	1	0	0	1	0	0	0
26	b7b7764cf	tell draw pl	3	09	0	0	0	0	0	0	0	0	0	0
27	b7b7764cf	found bunr	3	19	0	0	0	2	1	2	1	2	0	1
28	b7b7764cf	3	4	23	0	0	0	0	0	0	0	0	0	0
29	b7b7764cf	insane	4	21	1	0	0	1	0	1	0	0	0	0
30	b7b7764cf	wants slee	4	01	0	0	0	0	0	0	0	0	0	0
31	b7b7764cf	really hate	4	03	2	1	2	2	1	2	1	2	1	1
32	b7b7764cf	watch mat	3	04	0	2	0	1	-1	0	0	0	0	0
33	b7b7764cf	loved 9 alle	4	02	0	0	0	0	1	0	2	0	0	0
34	b7b7764cf	not sit futo	4	01	-1	0	-1	0	0	0	0	-1	0	0
35	b7b7764cf	spend less	3	06	0	1	0	0	0	0	2	0	0	1

Fig. 4 Calculating emotions

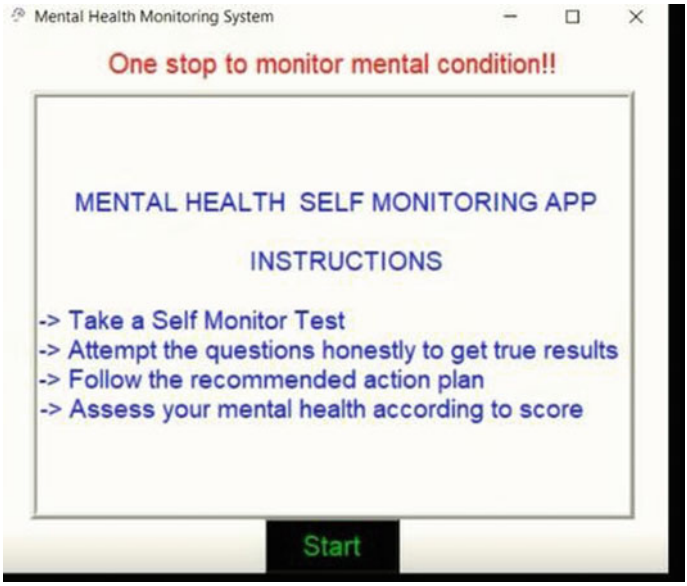

Fig. 5 Quiz analysis

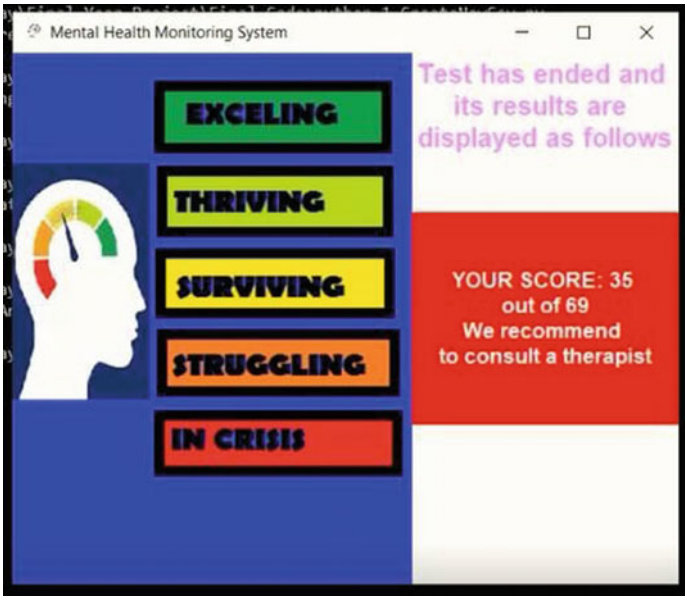

Fig. 6 Quiz analysis score

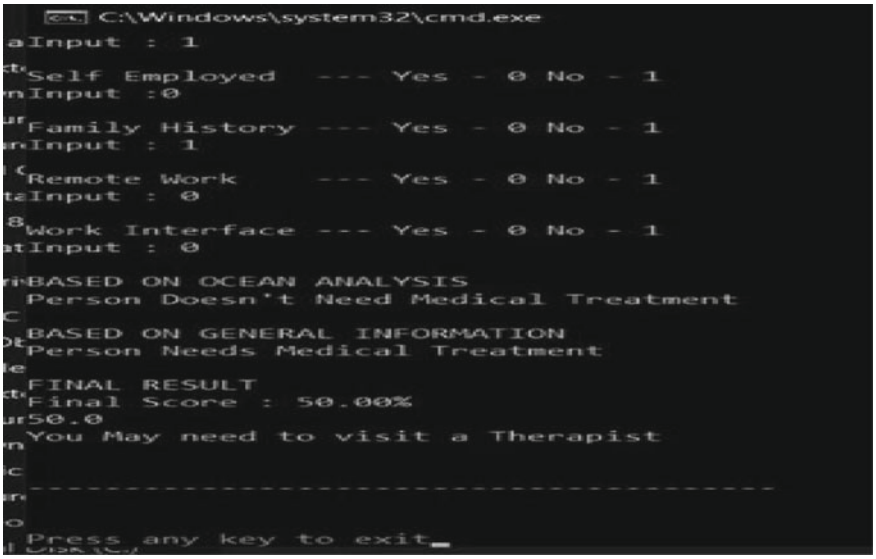

Fig. 7 Personal details

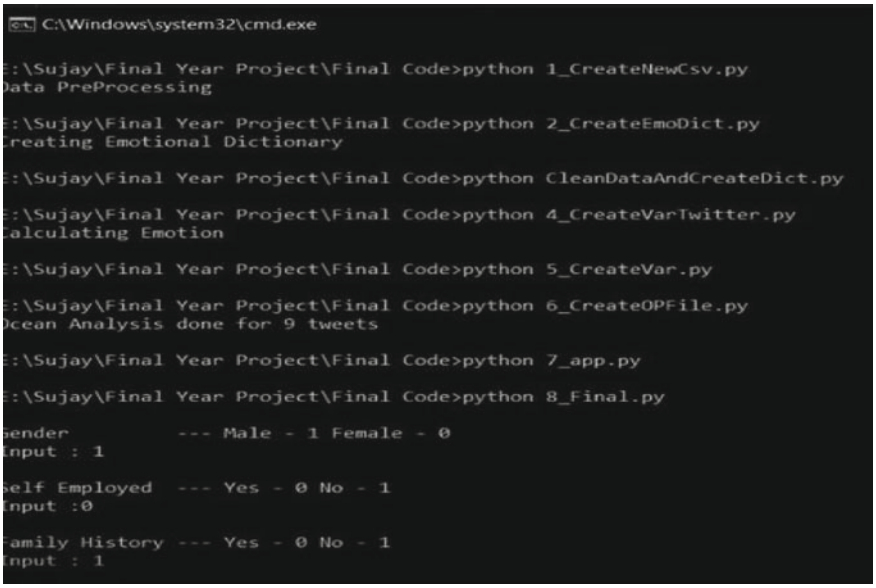

Fig. 8 Mental health prediction

References

1. K. Young, M. Pistner, J. O'Mara, J. Buchanan, Cyberpsychol. Behav. **2**(5), 475–479 (1999). https://doi.org/10.1089/cpb.1999.2.475 (PMID: 19178220)
2. Y. Mehta, N. Majumder, A. Gelbukh, E. Cambria, Recent trends in deep learning based personality detection. Artif. Intell. Rev. **53**, 2313–2339 (2020). https://doi.org/10.1007/s10462-019-09770-z
3. D. Xue, Z. Hong, S. Guo, L. Gao, L. Wu, J. Zheng, N. Zhao, Personality recognition on social media with label distribution learning. IEEE Access. https://doi.org/10.1109/ACCESS.2017.2719018
4. J. Block, Issues of DSM-V: internet addiction. Am. J. Psychiatr. **165**(3), 306–307 (2008). https://doi.org/10.1176/appi.ajp.2007.07101556 (PMID: 18316427)
5. K.S. Young, Internet addiction: the emergence of a new clinical disorder. Cyber Psychol. Behav. **1**, 237–244 (1998). https://doi.org/10.1089/cpb.1998.1.237
6. 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, C.-F. Yen, The association between suicidality and Internet addiction and activities in Taiwanese adolescents. Compr. Psychiat. (2014)
7. Y. Baek, Y. Bae, H. Jang, Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. Cyberpsychol. Behav. Soc. Netw. (2013)
8. D. La Barbera, F. La Paglia, R. Valsavoia, Social network and addiction. Cyberpsychol. Behav. (2009)
9. K. Chak, L. Leung, Shyness and locus of control as predictors of internet addiction and internet use. Cyberpsychol. Behav. (2004)
10. K. Caballero, R. Akella, Dynamically modeling patients health state from electronic medical records a time series approach. KDD (2016)
11. L. Zhao, J. Ye, F. Chen, C.-T. Lu, N. Ramakrishnan, Hierarchical Incomplete multi-source feature learning for Spatiotemporal Event Forecasting. KDD (2016)
12. E. Baumer, P. Adams, V. Khovanskaya, T. Liao, M. Smith, V. Sosik, K. Williams, Limiting, leaving, and (re)lapsing: an exploration of Facebook non-use practices and experiences. CHI (2013)
13. S.E. Jordan, S.E. Hovet, I.C.-H. Fung, H. Liang, K.-W. Fu, Z.T.H. Tse, *Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response*. Big Data and Digital Health
14. E. Heiervang, R. Goodman, Advantages and limitations of web-based surveys: evidence from a child mental health survey. Soc. Psychiat. Epidemiol. **46**, 69–76 (2011). https://doi.org/10.1007/s00127-009-0171-9