Detecting Early Risk of Depression
from Social Media User-generated Content

Hayda Almeida, Antoine Briand, and Marie-Jean Meurs

University of Quebec in Montreal (UQAM), Montreal, QC, Canada
`meurs.marie-jean@uqam.ca`

Abstract. This paper presents the systems developed by the UQAM team for the CLEF eRisk Pilot Task 2017. The goal was to predict as early as possible the risk of mental health issues from user-generated content in social media. Several approaches based on supervised learning and information retrieval methods were used to estimate the risk of depression for a user given the content of its posts in `reddit`. Among the five systems evaluated, the experiments show that combining information retrieval and machine learning approaches gives the best results.

Keywords: Information Retrieval, Mental Health, Natural Language Processing, Supervised Learning, Text Mining

1 Introduction

The Early Detection of Depression Pilot Task was part of the CLEF eRisk 2017 workshop [16]. The pilot task challenge consists of performing early risk detection of depression by analyzing user-generated content from `reddit`[1]. Towards this goal, a system receives user-generated content as input, and should output a prediction regarding the user's susceptibility to depression.

The pilot task dataset contains user-generated content, which is organized and processed chronologically. This allows for monitoring the user progress, and detecting risk as early as possible. Users are categorized as *risk* or *non-risk* (of depression). Each user produced a sequence of `reddit` posts, written within a given period of time.

The pilot task was organized in two stages: training and test, each having a different dataset divided into 10 chunks. During training stage, a dataset containing a sequential set of posts per user was provided along with the user's category. All training chunks were made available, containing the complete user post sequence.

During test stage, the dataset of test users was released sequentially (one release each week). Each release contained part of the user post sequence, corresponding to one chunk (from the oldest to the newest posts). Participant systems had to output predictions for users based on all current test chunks before the release of a new chunk. The predictions could be either the category of a user or no decision, up to the last week of the test stage where all the users had to be given a category.

---

[1] `https://www.reddit.com/`

We describe hereafter our prediction system based on an ensemble classification approach, which combines supervised learning, information retrieval, and feature selection methods. This report is organized as follows: the system resources are described in Section 3; the system modules, and the decision algorithm merging the module predictions are described in Section 4. Experiments and results are described in Section 5 while conclusions and future works are discussed in Section 6.

2   Related Work

Social media content has been commonly utilized to develop approaches that support mental health care. The latest CLPsych Shared Tasks [5,18] have proposed participants to predict users in eminent risk of depression or Post Traumatic Stress Disorder (PTSD). These tasks made use of tweets or mental health forum posts.

In [11], a sentiment analysis model was built with focus on user-generated social media content. It uses highly relevant sentiment lexicons and sentiment intensity measurements. The authors demonstrated that the approach outperforms other commonly used lexicons, as well as machine learning-based tools.

The authors of [19] evaluated the usage of different features to analyze user posts from `LiveJournal`[2], and compare discrepancies between posts from depression related online communities, and control (non-depression) related communities.

Another approach was proposed by [17], relying on a statistical model based on the analysis of over 176 million tweets to identify communication patterns related to mental illness in Twitter, and to attempt predicting user behavioral patterns related to depression. We describe hereafter studies conducted mainly based on two research fields: supervised learning, and information retrieval.

2.1   Supervised Learning for Mental Health

Several studies were conducted towards identifying mental health issues in social media by using supervised learning methods. The choice of supervised algorithms varies according to the tasks and data at hand. However, the previous studies presented here generally rely on highly discriminative features to achieve state-of-the-art performance. This demonstrates the importance of attribute choice for such tasks.

In [8], the authors presented a study on predicting depression from tweets by analyzing over 2 million posts of 476 users. The best performance was obtained with a SVM classifier and a set of behavioral features, such as occurrence of pronouns, usage of swearing and depression terms, tweet replies, as well as posting time and frequency.

The work presented in [14] identifies user psychological stress in tweets. Features such as emotion words, smileys, tweet mentions, replies, and posting frequency were obtained from single tweets, and from all user's tweets. The best performance was obtained by a four layer Deep Neural Network (DNN).

Previous works have also used Twitter data to identify language differences between users potentially presenting PTSD [6], or who attempted suicide [7]. In both these studies, the authors evaluated user-generated content using word and character language

---

[2] `http://www.livejournal.com`

models. The findings point to characteristics of tweets associated to mental health issues, such as heavier use of emotions, usage of third person pronouns, anxiety terms, as well as high posting frequency.

The authors in [23] analyzed Facebook [3] status updates to predict user satisfaction with life. Their approach used feature selection of n-grams and topic extraction, aand built regression models based on the message level, and the user level. The results indicate that a cascade model, using message level predictions to inform user level predictions, performed best.

2.2 Information Retrieval for Mental Health

Information retrieval techniques are widely used to support knowledge discovery in the biomedical field. Most of the approaches are designed to help researchers and practitioners looking for relevant documents to support experiments or diagnoses.

In the field of mental health, [10] reports an interesting study to support mental health maintenance of U.S. army soldiers. The goal is to aid health practitioners to perform efficient follow-ups on soldiers, since the suicide attempt rate among them is known to be high.

The approach made use of the Veterans Informatics and Computing Infrastructure (VINCI) resource to process mostly unstructured health information, such as clinical notes. The authors built a search engine based on Apache Solr[4] indexing these textual data to predict the risk of suicide attempt among soldiers. Even though only few pre-processing steps were utilized in this system, it provides promising performance, and covers a larger population than systems based on structured data.

3 Resources

The following Sections describe the resources utilized to build our systems.

3.1 Dictionaries

The supervised learning-based systems rely on a set of depression-related dictionaries. The dictionary keywords are used to provide discriminative attributes for automatic classification. The dictionaries we utilized are lists of relevant feelings, medicine, drugs, and diseases, which are assumed to be related to depression.

The feeling dictionary is composed of feeling words used in mental status exams[5], and a conceptual feature map obtained from SenticNet [4]. The medicine dictionary lists antidepressant names or depression-related medicine, obtained from Wikipedia[6]. The disease dictionary is composed of depression-related disease names, from Wikipedia[7].

---

[3] https://www.facebook.com
[4] http://lucene.apache.org/solr/
[5] http://psychpage.com/learning/library/assess/feelings.html
[6] https://en.wikipedia.org/wiki/List_of_antidepressants
[7] https://en.wikipedia.org/wiki/Depression_(mood)

The drug dictionary contains a list of psychoactive drug names, such as hallucinogens, psychedelics, anxiolytics, and sedatives, also obtained from Wikipedia[8].

3.2 Open Source Software

Classification To support developing the supervised learning method in our system, we have utilized the open-source machine learning framework Weka [24][9]. The Weka framework provides standard implementations of several classification algorithms. It also provides modules to handle and process Attribute Relation Format Files (ARFF) files, which contain a matrix representation of the dataset in terms of instances versus features, allowing to easily perform feature selection.

Indexating The information retrieval method in our system relies on the open-source search platform Apache Solr. The Solr platform allows for building a search engine to perform full-text search in a document index. Both Solr search and index modules are built based on the Apache Lucene[10] library. A Solr index is designed based on a schema, which is composed of a set of fields that represent a document object. Several pre-processing steps are also available in Solr, which can be applied at indexing time and also at query time.

4 Methodology

To detect users in risk of developing depression, we have designed a multipronged approach that combines results obtained from both Information Retrieval (IR) and Supervised Learning (SL) based systems. The combination is performed by a decision algorithm.

In Section 4.1, we explain how we utilized the CLEF eRisk training and test datasets in our experiments. The IR-based systems are described in Section 4.2 while the SL-based systems are presented in Section 4.3. Details on the decision algorithm are presented in Section 4.4. Finally, we briefly describe how we performed experiments to determine the best configuration for our approach in Section 4.5.

4.1 Dataset

The CLEF eRisk training and test datasets are composed of user posts extracted from `reddit`. Both datasets are divided into a total of 10 chunks each, chronologically organized. Each chunk represents a sequence of writings for a given user in a period of time. Table 1 shows statistics on the eRisk 2017 pilot task datasets.

We have utilized the chronological aspect of the user writings when processing both training and test data. When processing the training data, we have computed the user posting frequency, which is further described in Section 4.3. When processing the test data, we have considered single chunk and multiple chunk predictions, as further explained below.

---

[8] https://en.wikipedia.org/wiki/Psychoactive_drug
[9] http://www.cs.waikato.ac.nz/ml/weka/citing.html
[10] https://lucene.apache.org/core/

	Training dataset	Test dataset
# users	486	401
# writings	294,817	236,371
# no-risk users	403	349
# risk users	83	52
# no-risk writings	263,966	217,665
# risk writings	30,851	18,706

Table 1. Statistics on the eRisk 2017 pilot task dataset

Training The training set was provided in its completeness at the beginning of the task. It has been manually annotated by experts. Users are categorized as either *risk* (depressed) or *non-risk* (non-depressed).

To identify the most suitable models for both IR and SL methods, we performed several experiments using the training data. We utilized the training data in two different ways: first, using cross-validation on the training chunks 1 to 10; second, using the training chunks 1 to 9 as training set, and the training chunk 10 as validation set.

Test The test set was provided gradually, being each test chunk released one week apart from the previous test chunk. Predictions on the test set were therefore provided weekly by our systems.

In order to output predictions in a given week, we have utilized the test data in two different ways: first, to obtain a list of predictions only considering the current test chunk; second, to obtain a list of predictions considering all test chunks released so far. Both list of predictions are taken into account when merging outputs from different models and systems.

4.2 Information Retrieval Based Systems

We used an approach based on IR to retrieve similar documents from a test document used as a query. The intuition is that using the full content of a user post as a query should allow a search engine to retrieve semantically similar documents (posts). In our context, the similar posts are retrieved from the training corpus where they are already labeled according to the risk/no-risk state of the user who wrote them. We built two search engines relying on two different indexes created from the eRisk training corpus with and without indexing stop-words. We then considered the eRisk test documents as queries, which were submitted to both search engines.

For each test document d submitted to the search engines, we used the class (*risk* or *non-risk*) of the top n retrieved documents to compute a score $S_{IR}(d)$ reflecting how likely d has been produced by a depressed user. This can be compared to a *k-nearest neighbors* approach since we want to get the closest documents (neighbours) to a given

document. The number of retrieved documents taken into account has been set experimentally to $n = 20$. $S_{IR}(d)$ is computed as follows:

$$S_{IR}(d) = \frac{1}{n} \sum_{i=1}^{n} \delta(d_i)$$

where d_i is the document retrieved by query d in position i, and

$$\delta(d_i) = \begin{cases} 1, & \text{if } d_i \text{ is labeled as } \textit{risk} \\ 0, & \text{otherwise} \end{cases}$$

The test documents are then ordered according to their S_{IR} score, and considered as *risk* candidates if their score is above a given threshold, which was experimentally set. The search engines created in this approach rely on Apache Solr, and the BM25 probabilistic ranking algorithm [12]. We first indexed all the fields in the training set. Two indexes, I and II, were generated based on the same schema but applying different pre-processing steps, which are described in Tables 2 and 3.

For Index I, we indexed all the data with little pre-processing. Index II uses the same schema along with more pre-processing steps: stop-words removal, stemming (using the Solr built-in Porter Stemmer algorithm), and punctuation filtering.

Index name	Pre-processing
Index I	Tokenization
	Lowercasing
Index II	As Index I +
	Stemming
	Stopwords
	Punctuation

Table 2. Pre-processing steps by indexes

Table 3 presents the fields used in the schema, *i.e.* all the fields available in the corpus (title, content, date, label). The *Text* field is a copy field that contains both content and title, and is used as the default search field.

For better handling document-based queries, we utilized the built-in Solr MoreLikeThis (MLT) component [11]. Solr MLT enables retrieving documents that are similar to a given document, and is far more efficient compared to other classical search endpoints.

---

[11] https://cwiki.apache.org/confluence/display/solr/Other+
Parsers#OtherParsers-MoreLikeThisQueryParser

#	Indexed fields
1	Writing title
2	Writing content
3	Writing date
4	User label
5	Text (fields $1 + 2$)

Table 3. Indexed fields

4.3 Supervised Learning Method

The SL-based approach is based on the combined predictions of several classification models with different configurations. The SL models are designed using four classification algorithms and various feature types described below.

Features To design models for the SL-based systems, we have extracted discriminative features from the pilot task training dataset. Before extracting features, pre-processing steps were performed. These include word stemming, and normalization of URLs, smiley characters, as well as punctuation. The URLs and smiley normalization are relevant to better process the user-generated content, and help portraying the sentiment associated with a post. URLs can contain picture names, or words that refer to specific subjects. Smiley symbols are often used to represent an emotion, and during pre-processing they are replaced by actual words (e.g., `:)` or `:-)` are replaced by `happy`). All these cues are important since, if present, they might help representing a user's state of mind.

After pre-processing, four different feature types were extracted: n-grams, dictionary words, selected Part-Of-Speech (POS), and user posting frequency. N-gram features were extracted as of Bag-Of-Words (BOW), bigrams, and trigrams. Dictionary words were extracted based on the depression-related dictionaries described in Section 3.1. POS features were extracted by selecting the words annotated by the Stanford POS Tagger[12] as either adjective (JJ), noun (NN), predeterminer (PDT), particle (RP), or verb (VB).

As an attempt to account for the temporal evolution of the psychological state of a given user, we computed the user posting frequency, which represents the user activity pattern. The posting frequency of a user is computed as the time lapse between the oldest and the most recent writings, divided by the number of writings a user has generated in total. Statistics on features extracted from the training set are presented in Table 4.

Classifiers To build the SL models we have used three classification algorithms: Logistic Model Tree (LMT) [13], an Ensemble of Sequential Minimal Optimization (SMO) [20] (ens_SMO) classifiers, and an Ensemble of Random Forests [2] (ens_RF) classifiers.

---

[12] https://nlp.stanford.edu/software/tagger.shtml

	# Features
BOW	105,161
Bigrams	1,544,714
Trigrams	3,397,459
Selected POS	118,139
Feelings dic.	205
Medicine dic.	30
Drugs dic.	57
Diseases dic.	43

Table 4. Number of unique features

The ensembles are composed of 30 different classifiers each.
The 30 Random Forest classifiers composing the ens_RF were designed with iteration values from 10 to 50 (with increments of 10), and tree depth values from 2 to 10 (with increments of 2), as well as unlimited.
The 30 SMO classifiers composing the ens_SMO were designed with tolerance parameter values from 0.001 to 0.005 (with increments of 0.001), and epsilon for round-off error values from 1 to 5 (with increments of 1).

4.4 Decision Algorithm

The decision algorithm merges the predictions from both IR and SL based systems. The IR-based candidates are ranked based on similarity, and each candidate is associated with a S_{IR} score, as described in Section 4.2. Documents with highest scores are considered as candidates for the *risk* class. For the eRisk task, the high score threshold has been experimentally set to 0.7, *i.e.* all the candidates are documents d with a score $S_{IR}(d)$ such that $S_{IR}(d) \geq 0.7$.
The SL-based approaches are used to refine the list of candidates proposed by the IR-based systems. To be selected, a document from the IR-based list must be classified as *risk* by at least one of the SL-based systems. Candidates proposed by the SL-based system are also ordered according to the confidence of the prediction, and first ranked candidates are selected regardless of their presence in the IR list. The decision function Δ can be formalized as follows:

$$\Delta(d) = \mathbb{1}_{IR}(d) + \mathbb{1}_{SL}(d) + \mathbb{1}_{SLf}(d)$$

where d is a test document, and $\mathbb{1}_{IR}$, $\mathbb{1}_{SL}$, $\mathbb{1}_{SLf}$ are the indicator function respectively associated to the IR-based, the SL-based, and the SL-ranked-first lists of candidates. If $\Delta(d) \geq 2$, the document d is assigned the *risk* class, *i.e.* the user who generated this content is susceptible to depression.

4.5 Experiments

In order to determine the most suitable configuration for the IR and SL based systems, as well as the threshold for the decision algorithm, we have performed several experiments utilizing the pilot task training data.

The classification models were selected after performing experiments with all three classifiers using all feature types, or several feature types combined. Only the best performing combination of feature sets and classifiers were kept for the SL-based systems. For the experimental evaluation, the pilot task training dataset was utilized as described in Section 4.1.

The IR-based systems presented in Section 4.2 rank the users (writings) based on the $S_{IR}(d)$ score. This score is based on the categories of the 20 top similar documents retrieved. The number of documents in the top list has been setup through experiments on the training set. We ran several tests with different values (from 5 to 50, with increment of 5), and we chose 20 since it maximized the F-measure.

5 Results and Discussion

We submitted predictions on the test dataset obtained by five different systems. Four of these systems rely on a different ensemble configuration. The ensembles are either a merge of results obtained from the SL and IR based systems, or from a group of SL classifiers or IR-based systems. The five presented systems are described here:

– *UQAMA* is based on an ensemble approach, merging the output candidates from all SL-based systems (considering three classifiers and all features), with the output candidates from the IR-based systems.
– *UQAMB* is based on candidates proposed by both IR-based systems only. We considered *UQAMB* as our baseline system.
– *UQAMC* is based on SL models built with a LMT classifier, and using as features either BOW or bigrams separately, and BOW or bigrams together with all the dictionary features.
– *UQAMD* is based on SL models built with an ens_RF classifier, using as features either BOW or bigrams together with all the dictionary features.
– Lastly, *UQAME* is based on SL models built with an ens_SMO classifier, using bigrams separately and together with all the dictionary features.

The user posting frequency was a feature used by all five systems.

Table 5 present the results obtained by the five systems in terms of the metrics utilized by the CLEF eRisk pilot task. Besides F1, Precision, and Recall, the pilot task also evaluated systems using the early risk detection error (ERDE) [15]. The EDRE metric accounts for the imbalance problem on automatic classification, which could bias some classifiers. Additionally it penalizes late risk detection using a specific cost function, considering only the true positive scores, which are related to only the relevant (risk) documents.

In total, 8 teams participated in the CLEF eRisk 2017 pilot task, submitting a total of 30 different systems [16]. In Table 5, we highlight in bold the most interesting results

	ERDE 5	**ERDE 50**	**F1**	**P**	**R**
UQAMA	14.03%	12.29%	**0.53**	0.48	**0.60**
UQAMB	13.78%	12.78%	0.48	0.49	0.46
UQAMC	13.58%	12.83%	0.42	0.50	0.37
UQAMD	**13.23%**	**11.98%**	0.38	**0.64**	0.27
UQAME	13.68%	12.68%	0.39	0.45	0.35

Table 5. Performance results on the eRisk test set

obtained by our systems. Among our five presented systems, the best overall performance was achieved by *UQAMA* with the best F1 score and Recall. The best Precision was achieved by *UQAMD*, which is designed based on an ens_RF classifier. The contribution of each method to the performance of *UQAMA* needs to be further evaluated, as well as the impact of the various experimental settings.

Finally, an interesting observation was drawn from analyzing the user posts of candidates predicted as risk by our systems. The post content of such candidates often presented two major topics: "video games", and "sexuality or relationship issues". The relationship between "depression" and these two topics has been studied from a clinical perspective in several recent works [21,3,9,22]. Interestingly, the co-occurrence of these topics with risk of depression was also spotted by our systems.

6 Conclusion

This report describes the early risk prediction systems submitted to the CLEF eRisk 2017 pilot task. The system that performed best is based on a multipronged approach, which combines predictions from SL and IR based systems. SL-based systems made use of four major feature types, and three classification algorithms, LMT, ensemble SMO and ensemble RF. IR-based systems utilize two indexes, and users are ranked according to a similarity score based on the BM25 ranking algorithm [12].

The predictions obtained from both SL and IR based systems are merged by a decision algorithm. The results demonstrate that combining SL and IR approaches outperforms the results obtained by each approach applied separately.

Future work During our experimental phase, we have performed preliminary tests to evaluate the usage of three other methods: (1) simple rule-based classification using a sentiment analysis library, (2) deep learning-based classification using a Recurrent Neural Network (RNN), and (3) topic extraction using Latent Dirichlet Allocation [1]. Improving the system performance will involve further investigation of these approaches, as well as enhancement of the IR-based resources of the system.

Reproducibility Our system is publicly released as an open source software, and can be accessed at: `https://github.com/BigMiners/eRisk2017`

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3(Jan), 993–1022 (2003)
2. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
3. Brunborg, G.S., Mentzoni, R.A., Frøyland, L.R.: Is video gaming, or video game addiction, associated with depression, academic achievement, heavy episodic drinking, or conduct problems? Journal of Behavioral Addictions 3(1), 27–32 (2014)
4. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. pp. 1515–1521. AAAI Press (2014)
5. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M.: CLPsych 2015 shared task: Depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology (CLPsych): From Linguistic Signal to Clinical Reality. pp. 31–39 (2015)
6. Coppersmith, G., Harman, C., Dredze, M.: Measuring Post Traumatic Stress Disorder in Twitter. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM) (June 2014)
7. Coppersmith, G., Ngo, K., Leary, R., Wood, A.: Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the 3rd Workshop on Computational Lingusitics and Clinical Psychology (CLPSych). pp. 106–117 (2016)
8. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting Depression via Social Media. In: Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM). p. 2 (2013)
9. Granic, I., Lobel, A., Engels, R.C.: The benefits of playing video games. American Psychologist 69(1), 66 (2014)
10. Hammond, K.W., Laundry, R.J., OLeary, T.M., Jones, W.P.: Use of text search to effectively identify lifetime prevalence of suicide attempts among veterans. In: Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS). pp. 2676–2683. IEEE (2013)
11. Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM) (June 2014)
12. Jones, K.S., Walker, S., Robertson, S.E.: A probabilistic model of information retrieval: development and comparative experiments: Part 2. Information Processing & Management 36(6), 809–840 (2000)
13. Landwehr, N., Hall, M., Frank, E.: Logistic model trees. Machine Learning 59(1-2), 161–205 (2005)
14. Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L., Feng, L.: User-level psychological stress detection from social media using deep neural network. In: Proceedings of the 22nd ACM International Conference on Multimedia. pp. 507–516. ACM (2014)
15. Losada, D.E., Crestani, F.: A Test Collection for Research on Depression and Language Use. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 28–39. Springer (2016)
16. Losada, D.E., Crestani, F., Parapar, J.: eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental foundations. In: Proceedings Conference and Labs of the Evaluation Forum CLEF 2017. Dublin, Ireland (2017)
17. McClellan, C., Ali, M.M., Mutter, R., Kroutil, L., Landwehr, J.: Using social media to monitor mental health discussions- evidence from twitter. Journal of the American Medical Informatics Association (JAMIA) p. ocw133 (2016)

18. Milne, D.N., Pink, G., Hachey, B., Calvo, R.A.: CLPsych 2016 Shared Task: Triaging content in online peer-support forums. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology (CLPsych). pp. 118–127 (2016)
19. Nguyen, T., Phung, D., Dao, B., Venkatesh, S., Berk, M.: Affective and content analysis of online depression communities. IEEE Transactions on Affective Computing 5(3), 217–226 (2014)
20. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines. Tech. Rep. MSR-TR-98-14, Microsoft (April 1998)
21. Ramrakha, S., Paul, C., Bell, M.L., Dickson, N., Moffitt, T.E., Caspi, A.: The relationship between multiple sex partners and anxiety, depression, and substance dependence disorders: a cohort study. Archives of Sexual Behavior 42(5), 863–872 (2013)
22. Schou Andreassen, C., Billieux, J., Griffiths, M.D., Kuss, D.J., Demetrovics, Z., Mazzoni, E., Pallesen, S.: The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychology of Addictive Behaviors 30(2), 252 (2016)
23. Schwartz, H.A., Sap, M., Kern, M.L., Eichstaedt, J.C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D., et al.: Predicting individual well-being through the language of social media. In: Pacific Symposium on Biocomputing (PSB). vol. 21, pp. 516–527 (January 2016)
24. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: The WEKA Workbench. Online Appendix for "Data Mining: Practical machine learning tools and techniques". Morgan Kaufmann, 4 edn. (2016)