Detecting Addiction, Anxiety, and Depression by Users Psychometric Profiles

Anna Monreale
CS Dept., University of Pisa
Pisa, Italy
anna.monreale@unipi.it

Benedetta Iavarone
Scuola Normale Superiore
Pisa, Italy
benedetta.iavarone@sns.it

Elena Rossetto
CS Dept., University of Pisa
Pisa, Italy
e.rossetto@studenti.unipi.it

Andrea Beretta
ISTI-CNR
Pisa, Italy
andrea.beretta@isti.cnr.it

ABSTRACT

Detecting and characterizing people with mental disorders is an important task that could help the work of different healthcare professionals. Sometimes, a diagnosis for specific mental disorders requires a long time, possibly causing problems because being diagnosed can give access to support groups, treatment programs, and medications that might help the patients. In this paper, we study the problem of exploiting supervised learning approaches, based on users' psychometric profiles extracted from Reddit posts, to detect users dealing with *Addiction*, *Anxiety*, and *Depression* disorders. The empirical evaluation shows an excellent predictive power of the psychometric profile and that features capturing the post's content are more effective for the classification task than features describing the user writing style. We achieve an accuracy of 96% using the entire psychometric profile and an accuracy of 95% when we exclude from the user profile linguistic features.

CCS CONCEPTS

• **Human-centered computing** → **Social media**; • **Computing methodologies** → **Supervised learning by classification**; • **Applied computing** → Psychology.

KEYWORDS

mental disorders, psychometric profile, machine learning, Reddit, depression, anxiety, addictions

ACM Reference Format:
Anna Monreale, Benedetta Iavarone, Elena Rossetto, and Andrea Beretta. 2022. Detecting Addiction, Anxiety, and Depression by Users Psychometric Profiles. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3487553.3524918

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524918

1 INTRODUCTION

Mental, neurological, and behavioral disorders and substance use are common to all countries, and according to estimations, hundreds of millions of people worldwide are affected by such issues. Unfortunately, most of these disorders are neither diagnosed nor treated, negatively impacting the lives of individuals suffering from them. Indeed, the absence of measures against these disorders often exposes many people to experience isolation and job loss. Another critical issue related to these mental disorders is the social stigma that leads many affected individuals to search for support or information for a self-diagnosis in informal resources and platforms, such as blogs, forums, social media, instead of appropriate professional assistance. Here, individuals share their experiences, emotions, and personal facts by writing text messages without feeling judged. Thus, social media accumulate many data hiding individual psychological aspects. The possibility to analyze these textual messages to extract and study psychological aspects is an excellent opportunity for researchers because they represent a valuable alternative source instead of classical clinical texts [6].

Within the last decade, different studies have analyzed the use of language in social media and have correlated it with psychological phenomena, disorders, and individual personality [3, 15, 16, 19–21]. The large amount of data coming from online platforms like Reddit, Facebook, and Twitter led researchers to propose methods for detecting and characterizing different mental disorders such as depression and anxiety, based on Natural Language Processing (NLP) and text classification techniques. Most approaches developed in the past aimed to distinguish "neutral" users (i.e., users who never discuss mental health-related issues on the selected online platform) from users affected by a specific disorder (i.e., users discussing topics about that disorder), classifying a single post. These approaches are indeed valuable for extracting insights on the aspects of mental illnesses shared by different users but fail to understand the mental state of specific users.

In this paper, we propose a preliminary approach to study the content of posts written by the same user in a specific mental health-related subreddit, focusing on the psychometric features that let discriminate between the different subreddits. We leverage the content of the user's post to understand the themes discussed in the subreddit. We verify if the written content (the semantic information encoded in a user's post) helps predict the subreddit the user is writing in. We develop an analytical framework based on supervised learning approaches for distinguishing users affected by *Addiction*, *Anxiety* and *Depression* disorders. After analyzing the post content of each category and the corresponding subreddits, we propose the evaluation of three classification models (i.e. Decision Tree, Support Vector Machine, and Random Forest) that we train by exploiting the psychometric profiles of users extracted from their set of posts written in Reddit. We empirically assess the predictive power of the three different classification models. All of them show very good performance in accurately classifying users of each disorder category. This empirically proves that the psychometric profile is a good proxy for the detection of the three mental disorders under analysis. Our study also shows that features capturing the content of the post are more effective for this classification task rather than features describing the user writing style. Indeed, we achieve an accuracy of 96% with the use of the entire psychometric profile, and an accuracy of 95% when we exclude from the user profile linguistic features, while the prediction performance significantly drops when considering only the writing style.

The rest of the paper is organized as follows. Section 2 discusses related work on the study of mental disorders by social media. In Section 3 we introduce the information about the Reddit dataset under analysis and we analyze the content of the posts of each category. Section 4 introduces the analytical methodology we propose involving also the definition of the psychometric profile of a user; while in Section 5 we present our empirical evaluation. Section 6 concludes the paper with a discussion on our findings and possible future work.

2 RELATED WORK

There is vast ongoing research that exploits user-generated content on social media to analyze mental health-related concerns. Most of the approaches in the literature aim to find the differences between posts in mental health-related communities and posts in other neutral communities, in which users discuss topics others than mental health-related ones. De Choudhury et al. [7] trained a classifier on linguistic features to detect depression from tweets. Preotiuc-Pietroi et al. [21] extracted features by LIWC, Latent Dirichlet Allocation (LDA) [4] and frequent 1-3 grams from Twitter data to understand the personality of users with post-traumatic stress disorders. In [25], the authors combined NLP techniques, LIWC features, LDA topic modeling, and various machine learning approaches to learn how to distinguish depression posts from neutral posts. They observed that depression posts are characterized by words related to preoccupation with the self, feelings, anger, hostility, negative emotions, and suicidal thoughts, and also by words related to interpersonal processes, to hopelessness and meaninglessness, to fatigue and low energy, or insomnia and hyperactivity, and words of negation. Standard posts, instead, are more about events happening in the past, social relations, and advice seeking.

Regarding anxiety disorders, in [24] the authors analyzed posts from anxiety-related subreddits (r/*HealthAnxiety*, r/*socialanxiety*, r/*PanicParty*, r/*Anxiety*) and compared them with neutral posts using vector embeddings, LDA topic modeling, N-gram language models, and LIWC features. Then, they combined these data representations and built binary classifiers capable of discriminating anxiety-related posts from neutral ones. They identified correlations between anxiety and specific LDA topics such as school, alcohol, and drugs. In [10] the authors compared anxious users' language when they write in anxiety support communities with the language the same users use in neutral communities. Moreover, they studied the differences in the use of language in neutral communities between these users and users who have not written in anxiety-related subreddits. Related to addictions, in [14] authors studied the user's activity in Forum77, a forum dedicated to addiction recovery. They identified three classes of posts written in the forum, each one corresponding to a phase of the recovery process: 1) using drugs, 2) withdrawal (or detoxification), and 3) the recovery. To understand if this forum is an effective detoxification aid, they studied the transitions of user posts towards these phases. Even if almost 50% of users relapsed during their path to recovery, the authors discovered that the phase of recovery is the most probable phase of the last post of the users. Moreover, they analyzed the content of posts and the corresponding responses received for each phase. Using LIWC, they discovered that different linguistic features characterize posts of each phase.

In [5], the authors analyzed first-person narratives on Reddit from the people that use opioids. In particular, they used posts from r/*OpiatesRecovery* and r/*Opiates* to discover changes in the social networks and daily lives among opioids users during the COVID-19 pandemic. They found that the pandemic led to more social support and mutual aid on Reddit and that the pandemic had mainly two effects on opioids users: a group of users exploited the suspension of social obligations to enter remission, while others experienced a forced withdrawal. To those that experienced withdrawal, the Reddit community appeared to be fundamental to help the subjects to enter remission.

To the best of our knowledge, the studies in the literature have mainly focused on developing solid approaches and models to detect mental illness from user-generated content. What needs further development is the study of the markers and features that help models identify the risk of mental illness. Recently, research attention has shifted towards the evolution of a user's state and writing, studying the history of posts of a user and the predictive markers that can be derived from it [11, 17, 22, 27, 28]. By looking at the posting history of a user, it is possible to understand how a user changes over time and recognize indicators for mental health risk prediction [13, 18]. Indeed, recent research has shown that the patterns of evolution of markers for mental disorders are a useful proxy for identifying users that suffer from mental illness [27]. Our research forms part of this scenario, focusing on identifying relevant features for discriminating the users writing in different mental health-related subreddits. To tackle this aspect, we also rely on the creation of a user's psychometric profile, which has already been applied to detect the shifting of users' online discourse from depressive to suicidal [9].

3 EVALUATION DATASET

The data used for our analysis are obtained from the Pushshift Reddit Platform [2] that collects dumps of all publicly available Reddit data. We selected seven subreddits that discuss topics related to mental health and addictions: *r/Anxiety*, *r/depression*, *r/Drugs*, *r/HealthAnxiety*, *r/socialanxiety*, *r/stopdrinking*, *r/SuicideWatch*.

	N. Posts	N. Users
r/Drugs	76,053	7,725
r/stopdrinking	141,925	10,843
Addiction	217,978	18,287
r/Anxiety	40,367	8,843
r/HealthAnxiety	10,622	2,092
r/socialanxiety	11,736	2,895
Anxiety	62,725	11,385
r/depression	116,210	14,623
r/SuicideWatch	40,548	7,991
Depression	156,758	16,442
Total	437,461	46,114

Table 1: Data remaining after cleaning and selection

From these subreddits, we extracted posts (without comments) covering the period from 1 January 2014 to 1 May 2020.

We grouped the subreddits into three categories, according to the issue that they cover: *Addiction*, *Anxiety*, and *Depression*. We merged the subreddits *r/stopdrinking* and *r/Drugs* into the category *Addiction*, because they are related to the use of various substances that can cause addiction and the related substance use disorders. We merged the subreddits *r/Anxiety*, *r/HealthAnxiety*, and *r/socialanxiety* into the category *Anxiety*, because they are all related to anxiety disorders. We merged the subreddits *r/depression* and *r/SuicideWatch* into the category *Depression*, since suicidal ideation or attempts are some of the symptoms that are common among depressive disorders. We then applied some cleaning and selection processing using the following criteria:

- keeping the posts written only by authors who wrote uniquely in subreddits we grouped in a single category. If a user wrote posts in subreddits that we grouped into different categories, we removed all that user's posts. The aim is to focus the analysis on users dealing with only one of the considered mental health categories of issues and study their language by supposing they are not struggling with any of the conditions considered in other categories;
- removing all the posts written by moderators;
- removing all the posts having no words in their body. These posts contain images or URLs and do not provide any useful textual information;
- removing all the posts whose content results deleted or removed by their authors;
- removing duplicated posts within the same subreddit and between different subreddits in case of reposts.

Since the goal is to construct a user profile based on the history of the posts written by the user, we selected only users with at least 3 posts in the considered Reddit categories. Table 1 describes the data remaining after this selection phase. Although the three categories do not present a relevant difference in terms of the number of users, the number of posts significantly differs. Especially, the *Anxiety* group contains a small number of posts if compared to the other categories.

3.1 User Overlapping Analysis

In Table 1, the total number of users in each category is lower than the sum of the number of users in the subreddits composing that category. The same is valid for the total number of users we selected. This mismatch is because we selected users who have written in a single mental disorder-related category but not necessarily in a single subreddit of that category. From Figure 1 we can understand the degree of overlapping between the subreddits. The overlap is more evident for users who wrote in the subreddits *r/depression* and *r/SuicideWatch*, *r/Anxiety* and *r/HealthAnxiety*, and *r/Anxiety* and *r/socialanxiety*. The overlap is minimum between users active in *r/HealthAnxiety* and *r/socialanxiety* and between users writing in *r/Drugs* and *r/stopdrinking*, who tend to write in just one of the two subreddits.

Users who wrote in more than one subreddit did not necessarily write the same quantity of posts in the various subreddits. For instance, some of them have written most of their posts in *r/depression* and just some posts in *r/SuicideWatch*, while others have done the opposite by writing all their posts in *r/SuicideWatch* except for very few posts in *r/depression*. To analyze this aspect, for each user, we computed the ratio between the number of posts they wrote in a specific subreddit over the total number of posts they wrote in the corresponding category. Then, we plotted the empirical complementary cumulative distribution function (complementary ECDF) of these ratios grouped by category, as reported in Figure 2. Note that the plot related to *Depression* is composed of two specular lines because the subreddits in this category are two. Therefore, if an user wrote N posts in the subreddits about *Depression*, of which M are written in *r/SuicideWatch*, this means that they wrote $N - M$ posts in *r/depression*. The same reasoning applies to the *Addiction* category. In *Anxiety*, the situation is slightly different because the subreddits taken into account are three. For each of these plots, we can observe users' activities through the subreddits:

- Users who wrote in subreddits of *Addiction* tend to write in a single subreddit, as shown from Venn diagrams in Figure 1. Given the users who wrote in *Addiction*, 40.7% of them wrote all their posts in *r/Drugs*, while 57.8% wrote all their posts in *r/stopdrinking*;
- Given the set of users who wrote in the category *Anxiety*, 77.7% of them wrote at least one post in the subreddit *r/Anxiety* thus, 22.3% of them did not write any post in this subreddit. We also have that 57% of users wrote all their posts uniquely in *r/Anxiety*; 7% of users wrote all the posts in *r/HealthAnxiety* while 18.4% wrote at least one post in this subreddit; finally, 14.8% of users wrote all their posts in *r/socialanxiety* and 25.4% of the users wrote at least one post in this subreddit.
- 11.0% of the users who wrote in subreddits we merged into *Depression* wrote all their posts exclusively in *r/SuicideWatch*, while only 51.4% of them wrote all their posts in *r/depression*. This also means that 48.6% of the users wrote at least one post in *r/SuicideWatch*.

3.2 Post Content Analysis

For each subreddit grouped by category, we analyzed the most common words. We report the corresponding word clouds, obtained by using WordCloud Python library [12], after removing stop words and lemmatizing each word using Gensim library [23].

Addiction. Figure 3 reports the two word clouds corresponding the subreddits *r/Drugs* and *r/stopdrinking*. In *r/Drugs* posts, there is a wide appearance of specific terms about drug use and their effects (drug, weed, acid, dose, pill, meth, coke, smoke, trip, LSD, effect) and about drinking (drink, alcohol). Among the most common words, there are also terms related to addressing problems, asking for help,

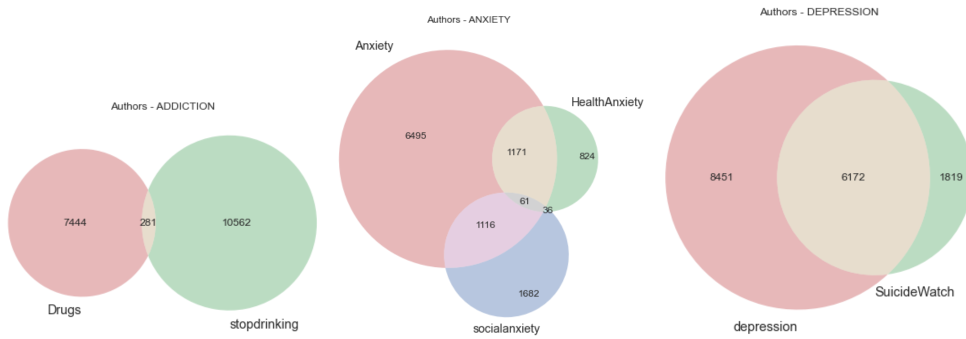

Figure 1: Users overlapping through subreddits

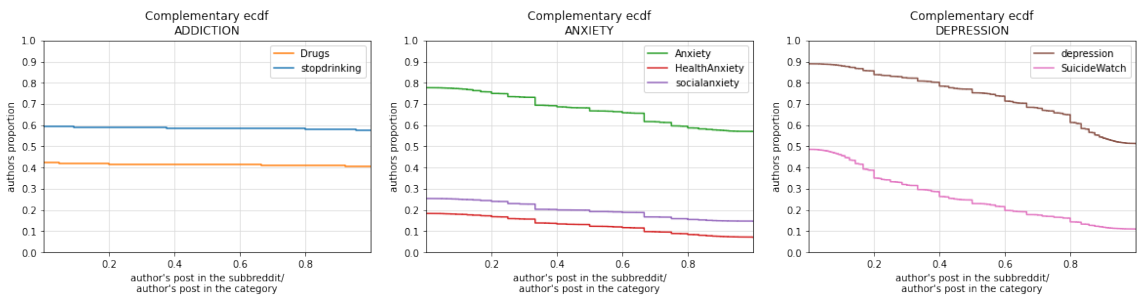

Figure 2: Authors activity through subreddits

Figure 3: Word clouds of *Addiction*

Figure 4: Word clouds of Depression

or information (problem, help, ask, need). In the word cloud are also visible terms about social relations (friend, person, experience, party) and personal concerns like work, school, home, and sleep. Terms related to anxiety are also among the most used ones (anxiety, Xanax). In *r/stopdrinking* the posts contain generic terms about alcohol and drinking (drink, drink today, alcohol, beer, alcoholic, bar, hangover) and terms about not drinking, stopping drinking, or the difficulties of stopping drinking (iwndwyt – acronymous of "I will not drink with you tonight", quit, end, recovery, quit drink, stay sober, stop drinking, sober, aa – Alcoholics Anonymous). Some positive terms appear in this subreddit, expressing hope and gratefulness (thank, proud, glad, hope, better). Common words are about working (work, job) and sleeping (sleep, bed), about family and social life (family, house, wife, husband, friend). Moreover, temporal references are often used (year, today, night, month, tonight, yesterday), and also in this subreddit, one of the most used words is anxiety.

Depression. Figure 4 reports the word clouds of the subreddits *r/depression* and *r/SuicideWatch*. Posts in *r/depression* contain mainly terms about depression, sadness, and loneliness (depress, depression, cry, lonely, sad, bad), problems, worries, fears, and uncertainty (don't know, afraid, think, thought, don't want, worry, maybe, scared, struggle, hard), death and pain (hurt, pain, cut, dead, want die, kill). However, many terms are also related to positive emotions (happy, fine, better, love), hope, search for help, and the users' wish to work on their mental condition (help, therapist, change, wish, try, hope). Terms regarding sleep, work, and money are frequent (bed, sleep, money, job) as social-related terms (high school, college, family, friend, parent, relationship, mother, girl, girlfriend, dad, best friend). Finally, differently from the above subreddits, "sorry" is one of the most common terms. It could be due to the sense of guilt, a common symptom of depressive disorders. In *r/SuicideWatch* some of the most used expressions are about suicide, death, and life (want die, suicidal, dead, suicidal thought, suicide, death, life, live, alive, want kill). Terms regarding sadness, depression, and pain (sad, depression, bad, pain, cry) are often used, along with words

Figure 5: Word clouds of Anxiety

about attempts and failures (attempt, try, fail, failure). Users in this subreddit use many negations (don't know, don't want, don't care, don't think, can't). Social-related terms are highly used (friend, girl, school, person, college, parent, relationship, dad, mother), and there are also terms about work, sleep, and money (job, sleep, money). Therapy and struggling with the situation are common themes (problem, struggle, deal, therapy). Also in *r/SuicideWatch* the term "sorry" is frequently used.

Anxiety. Figure 5 reports the word clouds of *r/HealthAnxiety*, *r/Anxiety*, and *r/socialanxiety*. In the subreddit *r/Anxiety*, terms related to anxiety and medical conditions are largely common (anxiety, anxious, panic attack, anxiety attack, stress, therapy, therapist, symptom, pain, medication, med, ill, doctor). Topics about work, school, and sleep are often discussed (work, job, school, sleep, tired). Some terms and expressions denote worries, struggles, insecurity, and fear (don't know, fear, worry, nervous, thought, think, don't want, afraid, die, problem, struggle, issue). Terms describing comparison (wrong, different, freak, normal) and terms about social and family life (relationship, parent, person, friend, family, kid, mom) are also present. In *r/HealthAnxiety* words anxiety-related and worries-related are frequently used. There are also common terms and expressions more specific to general health, illnesses or symptoms, and body parts. Users write in this subreddit to ask for advice. There are also terms related to family and friends, but with a lower frequency with respect to the subreddit *r/Anxiety*. In *r/socialanxiety* explicit references to anxiety are frequent (anxiety, anxious, social anxiety, sa – social anxiety, therapist). Here, expressions about interactions with others are more used (conversation, talk, relationship, date, speak, hang, tell). Users also write about people belonging to different social groups like family, friends, and coworkers (person, family, parent, friend, coworker, group, guy, girl) and about different social opportunities such as school (school, class, college, high school), work (work, job), and parties (party). There are copious common terms about what users feel and their struggle with their state (scared, uncomfortable, nervous, afraid, awkward, judge, fear, struggle, deal, avoid). These expressions are related to some typical symptoms of social anxiety, such as avoiding social circumstances, feeling judged, being embarrassed, feeling nervous, or somewhat scared in social situations [1]. Some frequent terms and expressions reflect the purpose of this subreddit to be a place where people can ask for help, seek advice, and share experiences (help, advice, question, problem).

Elements that are the most distinctive of each category are the following: the use of terms about drinking and specific lexicon about drugs in *Addiction*; wide use of explicit anxiety-related terms in *Anxiety*; negative emotions, references to death, and frequent use of negations in *Depression*. These considerations are consistent with the findings of De Choudhury et al. in [8].

3.3 Ethical considerations

User-generated data are often sensitive, and this is especially the case when the data are related to a topic like mental health. Despite treating delicate matters, the data we have used in this work are publicly available. The data were collected respecting Reddit Privacy Policy, in which users are made aware that the content they post on the platform, along with their username and other information, is indeed public and freely accessible by third parties (unless the users themselves delete the content or their account). We note that the risks associated with the data used in this work are very low. This work does not expose any usernames nor shows any complete sentences or posts written by a user, as it only reports words extracted from their context.

4 PREDICTING USER'S MENTAL HEALTH DISORDERS

To identify the user's mental health issues, we propose an analytical methodology that (*i*) extracts for each user their psychometric profile from the history of their Reddit posts and (*ii*) trains a machine learning model on the psychometric profile to predict the mental health issue.

Given the history of posts written by a user $P^u = p_1, \dots, p_n$, we extract their psychometric profile using LIWC[1] (Linguistic Inquiry and Word Count), a dictionary-based approach widely used in computational linguistics as a source of features for psychological and psycholinguistic analysis. LIWC Dictionary contains both style words (function words, articles, pronouns, prepositions, auxiliary verbs) and content words (conveying the content of communication). These two types of words have different roles, but both are interesting from a psychological point of view because it is important not only what people say but also how their communication is structured. This is also highlighted in [26] where Tausczik and Pennebaker state "style words are much more closely linked to measures of people's social and psychological worlds." LIWC processes the text in input, word by word, to obtain the final vector representation V, which is composed of 93 variables, i.e., $V = v_1, \dots, v_{93}$. The first feature v_1 represents the *word count*, i.e., the number of words in a post. Features $v_2, \dots v_8$ represent 7 summary variables summarizing aspects such as emotional tone or authenticity. Features $v_9, \dots v_{81}$ represent different aspects: Linguistic Dimensions and Grammar aspects, Psychological Processes (e.g., Affective, Social, Perceptual, Cognitive, Biological process). These variables

[1] Version LIWC2015

have a percentage value associated, i.e., the percentage of words in the text that belong to the corresponding LIWC category. The last set of features $v_{82}, \ldots, v_{93}$ corresponds to variables counting the punctuation marks in the text. In order to address the problem of distinguishing people with *Depression*, *Anxiety* and *Addiction* disorders, we consider the entire set of features of LIWC, except the features corresponding to the punctuation marks counts. In particular, given P^u, i.e., the list of posts of a user u, we extract from each post $p \in P^u$ the LIWC features V and then, we derive the user psychometric profile L^u as a vector of 81 features $L^u = l_1, \ldots, l_{81}$ computed as follows:

$$l_i = \begin{cases} \dfrac{\sum_{p \in P^u} f_i^p}{|P^u|}, & i = 1 \\ \dfrac{\sum_{p \in P^u} v_i^p f_1^p}{\sum_{p \in P^u} f_1^p}, & 2 \le i \le 81 \end{cases} \quad (1)$$

In the formula above, f_1^p denotes the words count in a post p. Thus, the first feature is the average of the word count with respect to the number of posts of the user, while the other features are obtained by the weighted average of each feature using the word counts of each post as weights.

To predict the mental disorder of a user, we employ classification approaches to estimate the likelihood of *Depression*, *Anxiety* and *Addiction* for a user based on the user profiles described in the previous section. The proposed framework is developed by using Decision Tree (DT), Support Vector Machine (SVM), and Random Forest classifiers (RF).

5 EXPERIMENTAL SETTING AND RESULTS

All the classification models (DT, RF, and SVM) were implemented in Python using the Scikit-learn library. We applied a training/test split selecting 70% (32,279 users) of data for training and 30% (13,835 users) for testing, performing a 5-fold cross-validation schema. Before building the classifiers, we removed all the features highly correlated, using the Pearson correlation index. Removed features are showed in Table 2. The performance of the classification techniques is evaluated with *Accuracy*, *Precision*, *Recall* and *F-measure* relying on a confusion matrix incorporating the information about each test sample prediction outcome.

Kept features	Removed features	Correlation
ppron	*pronoun*	0.88
i	*ppron*	0.83
ingest	*leisure*	0.83
negemo	*affect*	0.80
sexual	*swear*	0.80
time	*relativ*	0.79
anger	*swear*	0.79
function	*pronoun*	0.78
informal	*swear*	0.77
Analytic	*pronoun*	-0.81

Table 2: Highly correlated pairs of features

Table 3 reports the predictive performance of the three classification models tested in our framework. The results show that all classifiers perform very well, providing very high accuracy. More complex classifiers such as RF and SVM have better performance than DT, which instead has the advantage of being transparent by design. This means that even if the overall performance is low, a user can easily understand the reason for a specific classification.

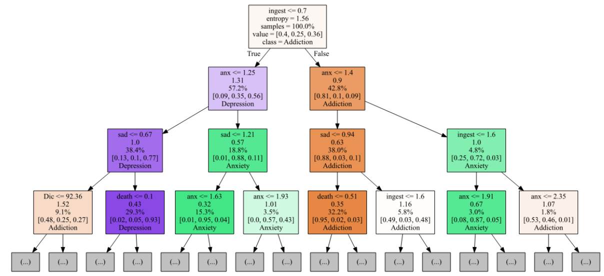

Figure 6: Split nodes of the first levels of the Decision Tree

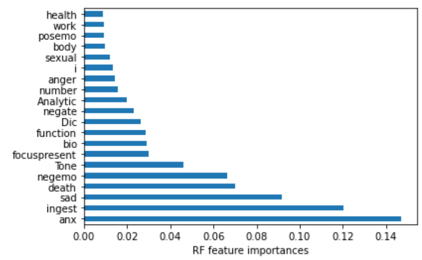

Figure 7: The 20 most important features of the psychometric profile for RF

However, we highlight that the DT obtained has a depth of 9 levels; thus, although easy to be comprehended, its complexity is not so low.

We report the visualization of the first levels of the DT in Figure 6 to support our discussion on the features that are important for the classification. We observe that the most important features of the psychometric profile used for classifying users are *ingest*, *anx* and *sad*, the only ones appearing in the split nodes of the first three levels. In the fourth level, to these categories we also have *death* and *Dic*.

Since the RF model is composed of several trees, it is not possible to analyze the structure of each of them, thus to understand which are the important features for the classification, we computed the feature importance that is reported in Figure 7 for the 20 most important features. The feature importance analysis shows that for the RF we have six features that stand out by importance from the others, and these are *anx*, *ingest*, *sad*, *death*, *negemo*, and *Tone*. We highlight that the first four features are the same identified in the first levels of the DT. We also observe that all the features related to the affective processes (*posemo*, *negemo*, *anx*, *anger*, *sad*) as well as all the features describing the biological processes (*bio*, *body*, *health*, *sexual*, *ingest*) are among the 20 most important features. The other features represent personal concerns (*death*, *work*) and time focus (*focuspresent*), or regard the linguistic dimension (*function*, *negate*, *i*). There are also some of the features related to the summary variables (*Dic*, *Tone*, *Analytic*).

For the SVM classifier, we analyzed the three hyperplanes to understand the influence of each feature on the classification. The results of this analysis are reported in Figure 8. Observing the three plots it is clear that *ingest*, *anx* and *sad* characterize the users of the three mental disorders.

Models	DT			RF			SVM		
Classes	Addiction	Anxiety	Depression	Addiction	Anxiety	Depression	Addiction	Anxiety	Depression
Precision	0.94	0.89	0.93	0.96	0.94	0.95	0.97	0.94	0.95
Recall	0.95	0.89	0.93	0.97	0.91	0.96	0.97	0.92	0.97
F-measure	0.94	0.89	0.93	0.97	0.92	0.95	0.97	0.93	0.96
Accuracy	0.93			0.95			0.96		

Table 3: Predictive performance of Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM) models.

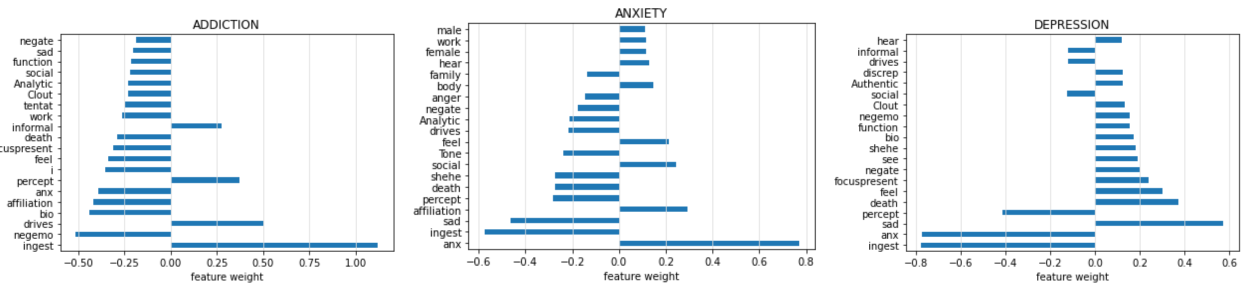

Figure 8: The most important features of SVM model. Each plot represents the 20 highest coefficients of the hyperplane dividing data points of a certain category from the others.

Concerning **Addiction**, among the highest positive coefficients of the hyperplane separating users of this class from the ones of the other two classes, there are those related to *ingest*, *drives*, *percept*, and *informal*. On the other hand, negative coefficients with higher absolute value include *negemo*, *bio*, *affiliation*, *anx*, *i*, *feel*, *focuspresent*, *death*, *work*, *tentat*, *Clout*, *Analytic*, *social*, *function*, *sad*, and *negate*. This indicates that users who are classified into *Addiction* use terms related to ingesting, to their goals and needs, and to their perceptions. Besides, they seem to use more informal terms than users in the other categories. Moreover, they also use fewer terms that express negative emotions, anxiety, or sadness, use fewer references to the first person singular, and are less focused on the present tense. Their posts follow a less analytical structure, and they do not exhibit a high social status or level of leadership. They also discuss less about death and work and use fewer function words and negations. Previous studies, such as [24][25], underline that people dealing with anxiety and depression tend to use more self-oriented references, and this is the only hyperplane identifying *i* among the most important features. The presence of a negative coefficient for *i* is in line with the findings of these studies since it means that users expressing more self-focus attention tend to be classified in *Anxiety* or *Depression*.

The hyperplane dividing authors of the **Anxiety** category from the others has the highest positive coefficients for *anx*, *affiliation*, *social*, *feel*, and *body*, *male*, *female*, *hear*, and *work*. Strongly negative coefficients are *ingest*, *sad*, *percept*, *death*, *shehe*, *Tone*, *Analytic*, *negate*, *anger*, and *family*. This indicates that, according to SVM, these users write more about anxiety and how they feel; moreover, their tone tends to be essentially negative, and they express less sadness or anger. Moreover, they write more about their needs of affiliation, social relations, work, and their bodies, but use fewer terms concerning ingesting and death, and they use fewer third-person singular references and negations. Their posts have a less analytical and more discursive structure.

Finally, the hyperplane separating users writing in **Depression** from the others is characterized by strongly negative coefficients corresponding to *ingest*, *anx*, and *percept*. This indicates that users of this class use fewer terms regarding ingesting, anxiety, and their overall perceptual processes. The highest positive coefficients are *sad*, *death*, *feel*, *focuspresent*, *negate*, *see*, *shehe*, *function*, *negemo*, *Clout*, and *Authentic*. Therefore, we can understand that these users express more sadness and negative emotions; they write more about death and their feelings, and their focus is mainly on the present. Moreover, they use more function words, their self-expression is more authentic and personal, and they use more references to the third person singular than users of the other categories.

A result shared by the different prediction models is that the most frequently misclassified users are the ones in *Anxiety* as the corresponding recall and precision values are the lowest for all the classifiers. This probably is due to the fact that this class contains a lower number of users and posts.

Set of features	LD and OG		Psychometric Profile without LD, OG, SV	
	Precision	Recall	Precision	Recall
Addiction	0.74	0.87	0.97	0.97
Anxiety	0.67	0.38	0.94	0.90
Depression	0.74	0.81	0.94	0.96
	F1-score = 0.69		*F1-score* = 0.95	
	Accuracy = 0.73		*Accuracy* = 0.95	

Table 4: Performance of SVM classifiers. On the left considering only features regarding "Linguistic Dimensions" (LD) and "'Other Grammar" (OG). On the right considering only all the features except for "Linguistic Dimensions" (LD), "Other Grammar" (OG) and "Summary Variables" (SV).

5.1 Influence of the writing style

The previous analysis highlights that the features related to LIWC dictionary categories which group mostly content words have greater importance to correctly predict the user mental disorder. Therefore, to analyze only the influence on the classification of the features describing the writing style, we performed an experiment where we removed from the psychometric profile of users all the information provided from the content of the posts written by users. Therefore, we built an SVM classifier (the best performer in the previous analysis) considering only features related to the parts of speech (POS), namely the ones related to "Linguistic Dimensions" (LD) and to "Other Grammar" (OG) of LIWC dictionary including elements as articles, auxiliary verbs, adverbs, conjunctions, impersonal and personal pronouns, negations, prepositions, verbs, comparisons, and interrogatives. The performance of the classification exploiting only this subset of features is reported in Table 4, on the left side. The results highlight that, using only the features describing the writing style, the predictive performance of the user's mental disorder drops. The users classified in the correct class are only 73% and, looking at the recall, more than 60% of users dealing with *Anxiety* are wrongly classified.

We also trained an SVM classifier on user profiles considering all the features except those corresponding to "Linguistic Dimensions", "Other Grammar" and "Summary Variables" of LIWC. Table 4 on the right reports the performance of this SVM model. Results show an increase in the accuracy, indeed now the corrected classified users are 95%, and the precision and recall become very good for all the classes. Comparing this performance against the SVM performance built on the full user profile (see Table 3), we can observe that they are almost the same. This means that even if the influence of features related to the writing style on the prediction is not null, it is limited.

5.2 Classification by important LIWC features

The results of the feature importance analysis presented above show that the most relevant category of LIWC features for the classification of *Addiction*, *Anxiety*, and *Depression* users is *affective processes*. However, although RF and SVM rank the various features differently, overall both consider important also features regarding also *biological processes* and some elements of *personal concerns*. Given this result, we tried to train an SVM model on the set of features belonging to these three LIWC categories and compare the results with an SVM built considering only the affective elements. The predictive performance of these classifiers is reported in Table 5. The results show that even if we established that the affective processes are key elements for our classification task, considering only these set of features to build a classifier, the accuracy is just 85%. This means that although the use of words related to sentiment and emotions is essential to discriminate between users dealing with *Addictions*, *Anxiety*, and *Depression*, affective processes features need to be considered together with features describing other aspects to improve the quality of the classification. The performance improves when we consider also features related to biological processes and personal concerns, reaching an accuracy of 94% (only two percentage points less than the accuracy of the classifier built on the entire set of features).

6 CONCLUSION

In this paper, we analyzed Reddit users dealing with three categories of mental disorders: anxiety, depression, and substance use (addictions). We modeled the users through a psychometric profile derivable from the posts they wrote in the categories. We showed how the psychometric profile could be an adequate proxy for predicting the user's mental disorders by training several supervised

Set of features	Affective & Biological Proc. + Personal Concerns		Affective Processes	
	Precision	Recall	Precision	Recall
Addiction	0.94	0.96	0.82	0.88
Anxiety	0.93	0.88	0.88	0.80
Depression	0.93	0.95	0.87	0.85
	F1-score = 0.93		*F1-score* = 0.85	
	Accuracy = 0.94		*Accuracy* = 0.85	

Table 5: Performance of SVM classifiers trained on features belonging to the LIWC categories: affective process, biological process and Personal Concerns.

models. Although the best predictor is an SVM model, all models have a good performance in solving this task, confirming the effectiveness of the psychometric profile. Our experiments also show that there are sub-sets of LIWC features that are more relevant and effective for our task. Features belonging to the *affective processes*, *biological processes* and *personal concerns* have a very good predictive power, while features describing the writing style have little influence on the detection of mental disorders. Another insight derived from our experimentation is that for any trained classifier, it results that the most frequently misclassified users are the ones dealing with *anxiety*.

Although the supervised approaches have optimal performance, there are limitations in our work that motivate future research and the extension of this study. First, the present study has only investigated features extracted by LIWC. Despite being widely used in literature and approaches similar to ours, these features might not be sufficient to capture complex phenomena such as mental disorders in all their facets. We believe this study would benefit from a broader set of NLP features, such as classic word embeddings or BERT-generated embeddings, that would capture contextual information, an aspect that LIWC features completely ignore. We intend to consider also LDA topic modeling, a technique that could give a better understanding of the themes discussed in the subreddits and let us further investigate the findings of this work. Another limitation is that this study was not specifically designed to consider subjects' mental evolution or role within the subreddit. In the classification, we assume that a user writing in a subreddit is affected by the mental disorder discussed and that the user is in a static role of seeking help, narrating a mental health experience, discussing their issues. This assumption ignores the variety of users who write on a subreddit, along with the fact that a user could shift from seeking help to giving help and support. The data might not include only subjects suffering from a disorder. A deeper investigation of the composition of the dataset is needed, possibly with rounds of manual annotation.

Further work will also concentrate on including other disorders or users writing in neutral subreddit in the data. In this way, we could verify the utility of the psychometric profile in detecting other forms of disorders and understand if the predictive performance changes in case we need to predict also users without a disorder.

ACKNOWLEDGMENTS

This work is partially supported by the European Community H2020 programme under the funding schemes: H2020-INFRAIA-2019-1: Research Infrastructure G.A. 871042 SoBigData++ (sobigdata.eu), G.A. 952215 TAILOR, G.A. 952026 Humane AI NET (humane-ai.eu).

REFERENCES

[1] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed. ed.). Autor, Washington, DC.

[2] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 830–839.

[3] Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multi-task learning for mental health using social media text. *arXiv preprint arXiv:1712.03538* (2017).

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.

[5] Amanda M Bunting, David Frank, Joshua Arshonsky, Marie A Bragg, Samuel R Friedman, and Noa Krawczyk. 2021. Socially-supportive norms and mutual aid of people who use opioids: An analysis of Reddit during the initial COVID-19 pandemic. *Drug and alcohol dependence* 222 (2021), 108672.

[6] Rafael Calvo, David Milne, Sazzad Hussain, and Helen Christensen. 2017. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering* 23 (01 2017), 1–37. https://doi.org/10.1017/S1351324916000383

[7] Munmun Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Conference on Human Factors in Computing Systems - Proc.* 3267–3276. https://doi.org/10.1145/2470654.2466447

[8] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8.

[9] Benedetta Iavarone and Anna Monreale. 2021. From depression to suicidal discourse on Reddit. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 437–445.

[10] Molly Ireland and Micah Iserman. 2018. Within and between-person differences in language used across anxiety support and neutral reddit communities. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. 182–193.

[11] Jina Kim, Jieon Lee, Eunil Park, and Jinyoung Han. 2020. A deep learning model for detecting mental illness from user content on social media. *Scientific reports* 10, 1 (2020), 1–6.

[12] Oesper L., Merico D., Isserlin R., and Bader G. D. 2011. WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source code for biology and medicine* 6, 1 (2011), 7.

[13] David E Losada, Fabio Crestani, and Javier Parapar. 2020. eRisk 2020: Self-harm and depression challenges. In *European Conference on Information Retrieval*. Springer, 557–563.

[14] Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An analysis of an online health forum dedicated to addiction recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1511–1526.

[15] Diego Maupomé and Marie-Jean Meurs. 2018. Using Topic Extraction on Social Media Content for the Early Detection of Depression. *CLEF (Working Notes)* 2125 (2018).

[16] Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk. 2014. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing* 5, 3 (2014), 217–226.

[17] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. 88–97.

[18] Javier Parapar, Patricia Martín-Rodilla, David E Losada, and Fabio Crestani. 2021. Overview of eRisk at CLEF 2021: Early Risk Prediction on the Internet (Extended Overview). (2021).

[19] Gregory Park, H. Schwartz, Johannes Eichstaedt, Margaret Kern, Michal Kosinski, David Stillwell, Lyle Ungar, and Martin Seligman. 2014. Automatic Personality Assessment Through Social Media Language. *Journal of personality and social psychology* 108 (11 2014). https://doi.org/10.1037/pspp0000020

[20] Sayanta Paul, Sree Kalyani Jandhyala, and Tanmay Basu. 2018. Early Detection of Signs of Anorexia and Depression Over Social Media using Effective Machine Learning Frameworks.. In *Proceedings of CLEF (Working notes)*.

[21] Daniel Preoţiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age, and gender in tweeting about mental illness. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*. 21–30.

[22] Andrew G Reece, Andrew J Reagan, Katharina LM Lix, Peter Sheridan Dodds, Christopher M Danforth, and Ellen J Langer. 2017. Forecasting the onset and course of mental illness with Twitter data. *Scientific reports* 7, 1 (2017), 1–11.

[23] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50.

[24] Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.

[25] Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2019. Detection of depression-related posts in reddit social media forum. *IEEE Access* 7 (2019), 44883–44893.

[26] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[27] Ana-Sabina Uban, Berta Chulvi, and Paolo Rosso. 2021. An emotion and cognitive based analysis of mental health disorders from social media data. *Future Generation Computer Systems* 124 (2021), 480–494.

[28] Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848* (2017).