

Life is not Always Depressing: Exploring the Happy Moments of People Diagnosed with Depression

Ana-Maria Bucur^{1,2}, Adrian Cosma³, Liviu P. Dinu¹

¹ University of Bucharest, Romania ² Universitat Politècnica de València, Spain

³ University Politehnica of Bucharest, Romania

ana-maria.bucur@drd.unibuc.ro, cosma.i.adrian@gmail.com, ldinu@fmi.unibuc.ro

Abstract

In this work, we explore the relationship between depression and manifestations of happiness in social media. While the majority of works surrounding depression focus on symptoms, psychological research shows that there is a strong link between seeking happiness and being diagnosed with depression. We make use of Positive-Unlabeled learning paradigm to automatically extract happy moments from social media posts of both controls and users diagnosed with depression, and qualitatively analyze them with linguistic tools such as LIWC and keyness information. We show that the life of depressed individuals is not always bleak, with positive events related to friends and family being more noteworthy to their lives compared to the more mundane happy events reported by control users.

Keywords: happiness, depression, social media, positive-unlabeled learning, LIWC, BERT

1. Introduction

“I have the choice of being constantly active and happy or introspectively passive and sad. Or I can go mad by ricocheting in between”, wrote the famous poet Sylvia Plath in her personal diaries (Plath and Hughes, 2000). It is now known that Sylvia Plath was clinically depressed most of her life, before tragically committing suicide. Depression is the modern ailment of our times. It is reported that more than 264 million people of all ages suffer from this mental illness ¹. More than 700 thousand people take their own life every year, and even more people attempt suicide ². Thankfully, modern psychology has a fair understanding of depression, its onset, its symptoms and potential treatments, more recently aided by computational methods (Orabi et al., 2018; Burdisso et al., 2019; Uban and Rosso, 2020; Inkpen et al., 2021; Bucur and Dinu, 2020; Bucur et al., 2021a). However, research in automatic understanding of manifestations of moments of happiness is lacking. How much do we understand about the happy moments of depressed individuals? In this preliminary work, we aim to bridge this gap and develop a computational method for extracting and analyzing happy moments from a large corpus of social media text.

Indeed, promoting happiness is an important aspect of our lives, and is one of the main ways in which we can prevent disease and live a longer, healthier life. While our happiness is determined largely by genetic factors (Diener et al., 1999), it can be influenced in a significant manner through our behavior: long-term relationships with friends and family, meaningful work and self-care. Paradoxically, Ford et al. (2014) showed that individuals in cultures that highly value happiness

are less likely to attain it. Moreover, the authors show that highly valuing happiness is a key indicator and a potential risk factor of depression. In the same vein, Alarcon et al. (2013) found that happiness is positively related to optimism and hope and that these traits are negatively associated with depression and anxiety. Recent computational tools (Caragea et al., 2018; Cobeli et al., 2022) currently allow for accurate modelling of the optimism-pessimism continuum from text, but very little research is performed on the specific instances of happy moments, and on what aspect makes an individual happy. We posit that while both depressed and control subjects experience happy moments, the reasons for their happiness differ.

In this preliminary work, we explore through linguistic tools the qualitative difference of happy moments between individuals diagnosed with depression and controls in social media text. While the majority of work in the area of exploring manifestations of depression in text focuses on specific symptoms (Inkpen et al., 2021; Bucur and Dinu, 2020; Bucur et al., 2021a), we focus our attention on exploring happiness manifestations in both depressed individuals and control. For this goal, we make use of HappyDB (Asai et al., 2018), a database of 100.000 expressions of happy moments, and conduct our analysis on eRisk (Losada and Crestani, 2016), a popular dataset of social media posts from Reddit for early detection of mental disorders. Different from sentiment polarity, which determines the sentimental aspect of an opinion, happiness is a broader concept that incorporates well-being and psychological health, referring to a positive affective state. Since there is no available dataset for modeling happy moments in depressed individuals and control, we use the Positive-Unlabeled learning paradigm to automatically extract happy moments from users’ posts, by combining HappyDB with eRisk to obtain a Positive-

¹<https://www.who.int/news-room/fact-sheets/detail/depression>

²<https://www.who.int/news-room/fact-sheets/detail/suicide>

Unlabeled dataset. We train an SVM in this paradigm (Elkan and Noto, 2008) on top of BERT (Devlin et al., 2019) embeddings to extract happy moments, and use LIWC (Pennebaker et al., 2001) and keyness statistics to make a qualitative exploration of the happy moments. This work paves the way to a more in depth understanding of depression through computational methods. In this way, automatic tools can be developed to aid in early diagnosis, therapeutic chatbots, and can provide more insights for developing suitable positive psychological interventions (PPIs) (Parks and Layous, 2016) (e.g., feeling and expressing gratitude, pursuing meaning, building hope) for people diagnosed with depression. PPIs have been shown to improve physical and mental health in patients with symptoms of depression (D’raven et al., 2015).

Through our analysis, we show that happy moments differ in people diagnosed with depression compared to control subjects, users from the depression group are more likely to consider happy moments those events related to family and friends. Our analysis shows that, for users from the control group, the main reasons for happiness are related to entertainment situations and financial plans.

This paper has the following contributions:

1. We propose a method for automatically extracting happy moments from social media text using the Positive-Unlabeled learning paradigm.
2. We make a qualitative exploration of happy moments present in the Reddit posts of both depressed individuals and controls and found noteworthy differences between the two groups, which are confirmed by psychological studies.

2. Related Work

In psychology, there are different scales for measuring the happiness of an individual, such as The Oxford Happiness Questionnaire (Hills and Argyle, 2002), The Satisfaction with Life Scale (Diener et al., 1985), The PNAS Scales (Watson et al., 1988) and others. The PNAS Scales measures different feelings or emotions (e.g., enthusiastic, alert, nervous, excited) in specific periods of time (e.g., in this moment, today, past few days), while The Oxford Happiness Questionnaire and The Satisfaction with Life Scale are comprised of different statements on happiness, unhappiness and life satisfaction.

In the field of computational linguistics, Mihalcea and Liu (2006) proposed a corpus-based approach to finding happiness. The paper analyzes a corpus of blog posts from the LiveJournal community annotated with sad and happy emotions. The authors’ findings reveal some patterns of happiness and sadness: 3 pm and 9-10 pm are the happiest hours, and Wednesday is the saddest day of the week. After analyzing the words from different semantic classes (e.g., actions, things, socialness), the authors conclude the work with a recipe of

happiness with the following ingredients: ”something new”, ”lots of food that you enjoy”, ”your favorite drink” and ”an interesting social place”.

In the realm of understanding manifestations of happiness, most notably, the Workshop on Affective Content Analysis proposed to tackle the problem of modeling happy moments in the CL-Aff Shared Task 2019 – in Pursuit of Happiness (Jaidka et al., 2019). It had two tasks: a) predicting the Agency (if the author is in control, if the happy moment is influenced entirely by their actions) and Sociality (if other people besides the author are involved in the happy moment) of happy moments and b) suggesting insights to characterize the happy moments automatically.

Even if many works exploring the social media discourse of people diagnosed with depression (Orabi et al., 2018; Burdisso et al., 2019; Uban and Rosso, 2020; Bucur et al., 2021c) are paying attention to the emotions expressed in their social media discourse (Aragon et al., 2021; Lara et al., 2021; Howes et al., 2014), to the best of our knowledge, works focusing on happiness felt by individuals diagnosed with depression are missing. Therefore, we aim to fill this gap and explore the happy moments from the online discourse of users with depression in comparison with control users.

3. Data

In our experiments, we use two datasets comprised of English texts: HappyDB and the eRisk 2018 depression dataset.

HappyDB (Asai et al., 2018) is a corpus of 100,000 happy moments, comprised of sentences with crowdworkers’ responses to the questions ”*what made you happy in the past 24 hours?*” or ”*what made you happy in the past 3 months?*”. It is a diverse corpus, containing happy moments from a variety of topics such as family (e.g., ”*My daughter waved at me and said ’mama’ for the first time when I came home this morning.*”), food (e.g., ”*I ate a juicy piece of pizza.*”), work (e.g., ”*It made me happy to get a promotion at work.*”), entertainment (e.g., ”*I went to a concert with friends.*”) and others.

In this paper, we use the cleaned version of the HappyDB, in which spelling errors were corrected and empty or single-word sentences were removed. An important aspect of HappyDB is that it contains only positive examples (i.e., statements of happiness) and does not contain any negative examples, which makes it difficult to use in standard classification settings.

The eRisk 2018 depression dataset (Losada and Crestani, 2016) contains posts from Reddit and it was used for the early risk detection of depression task from the Early risk prediction on the Internet (eRisk) Workshop³. Users were annotated as having depression by their mention of diagnosis (e.g., ”*I was diagnosed with depression*”) in their posts. Expression such as ”*I have depression*” or ”*I am depressed*” were not taken into

³<https://early.irlab.org/>

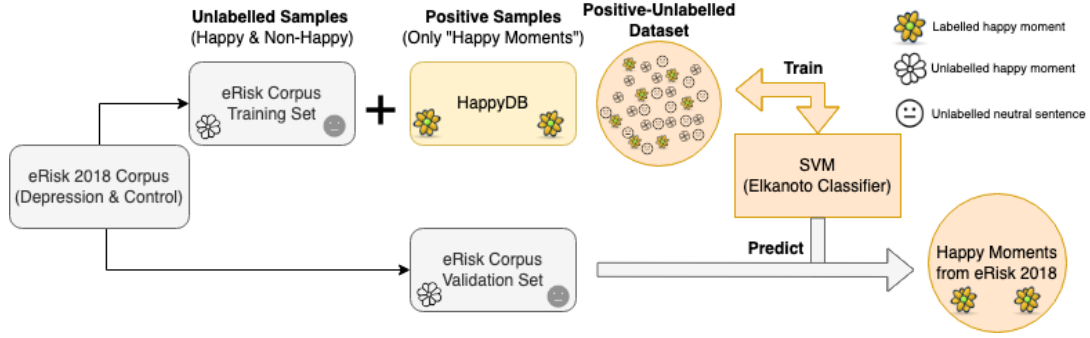


Figure 1: Positive-Unlabeled training paradigm, applied to extracting happy moments from eRisk depression dataset. We randomly sample 50,000 sentences from eRisk (unlabeled positive and negative classes) and 50,000 sentences from HappyDB (labeled positive class) to construct a Positive-Unlabeled dataset. After encoding the sentences with BERT, we trained a linear SVM in the PU-learning paradigm and use it to extract happy moments from eRisk validation set.

account in annotating the users, only users with explicit mentions of depression diagnosis were labeled as having depression. Users from the control group are random users who do not have any mention of diagnosis in their posts, some of them may be active in the depression subreddit⁴. The posts disclosing the depression diagnosis were removed by the authors from the dataset. The datasets contains 214 users diagnosed with depression and 1493 control users with approximately 90,000 submissions from the depression group and over 985,000 submissions from the control group.

4. Methods

To conduct our analysis, we first trained an SVM model on top of BERT sentence embeddings that is able to extract happy moments from social media text. We then analyzed the happy moments using LIWC (Pennebaker et al., 2001) and keyness information.

4.1. Extracting Happy Moments

In order to extract relevant happy moments from the eRisk dataset, we train a classifier according to the Positive-Unlabeled Learning framework (Elkan and Noto, 2008). In this setting, the goal is to train a binary classifier to distinguish between positive classes (i.e., the happy moments) and negative classes (i.e., any other type of text). However, during classifier training, only some of the positive examples are labeled, and none of the negative examples are labeled. Formally, we have a dataset of triples (x, y, s) , where x is an example, y is the class label, $y \in \{0, 1\}$, and s is a variable representing whether the pair (x, y) is labeled or not. As such, if the example is labeled, it belongs to the positive class: $P(y = 1 | s = 1) = 1$. When variable $s = 0$, the pair can belong to either class, 0 or 1 in our case. We chose to adopt the PU-learning methodology, as opposed to a strictly unsupervised approach, to include additional information in the computation from the eRisk dataset, which contains Reddit

posts qualitatively different from crowd-sourced texts from HappyDB.

In our case, we split the posts from the eRisk dataset into sentences, and construct a training dataset comprised of approximately 50,000 sentences from HappyDB (Asai et al., 2018) and 50,000 sentences from the posts of the first 10 subjects from eRisk (from both depression and control classes). This dataset is a Positive-Unlabeled dataset (PU dataset), since the sentences from HappyDB are always the positive class, but the sentences from eRisk can contain both happy moments and neutral (non-happy) text. This methodology is presented in Figure 1. Due to limited computational resources, we choose to use only a subset of the datasets for training our model. We use the rest of the subjects as a validation set to conduct our analysis.

We train a linear SVM (Cortes and Vapnik, 1995) with high regularization using the Elkanoto (Elkan and Noto, 2008) training methodology on the PU dataset. We extract semantic features using a pretrained SentenceBERT (Devlin et al., 2019) model and train the SVM on the sentence representations. We used the *pulearn*⁵ python library to train our classifier. To conduct our analysis, we used the previously trained SVM model on our eRisk validation set. Examples of extracted happy moments from eRisk validation set are presented in Table 1.

4.2. Exploring Happiness

For exploring the happiness moments expressed in the eRisk 2018 dataset, we use the Linguistic inquiry and word count (LIWC) lexicon (Pennebaker et al., 2001) due to its psychologically meaningful categories. The LIWC 2001 lexicon contains words from several categories such as linguistic dimensions (e.g., pronouns, articles, negations), psychological processes (e.g., affective processes, cognitive processes), and personal concerns (e.g. occupation, leisure). In our analysis, we aim to find the origin of people’s happy moments, thus

⁴<https://www.reddit.com/r/depression/>

⁵<https://github.com/pulearn/pulearn>

Depression	Control
Just found my first Rush dream setlist from long ago.	Yesterday morning, we woke up to a remarkably pleasant surprise.
I just ordered two pairs of these...they look super comfy!	I visited Japan recently.
I lost 120lbs through diet and exercise.	This new vegan supermarket just opened near me!
Getting under 200 is the best feeling!	I'm on vacation and there's this amazing yarn shop nearby!
Did get a novella published with a small press, working on bigger stuff.	I recently quit smoking weed.
I decided that I had to leave the job and make my new job the task of trying to improve my situation.	About a month ago, my dad finally went into alcohol rehab.
I recently adopted a cat and he's settled into his new home.	We went to some amazing gigs and hung out.
My sister bought me The Last of Us Remastered and I've just been grinding through it.	I get all excited when I receive an actual letter from someone I personally know.

Table 1: Selected happy moments automatically extracted from our validation set of eRisk dataset using Positive-Unlabeled learning. Different from sentiment polarity, happy moments express significant moments of positive affective states.

we analyze only the suitable categories from LIWC; we are not interested in the linguistics dimensions. As such, we chose the categories of sensory and perceptual processes (e.g. HEARING, FEELING), social processes (e.g., FAMILY, FRIENDS), and personal concerns (e.g., SCHOOL, SPORTS, RELIGION).

For comparing the happy moments from the two corpora (depression and control), we compute the dominance scores (Mihalcea and Pulman, 2009) for the LIWC features analyzed, thus comparing the relative difference in the use of words from different categories between users diagnosed with depression and control users. The coverage is computed by comparing the words from the foreground corpus (F) with the words from the background corpus (B). The coverage of a specific class (C) in a corpus (X) is computed as follows:

$$Coverage_X(C) = \frac{\sum_{w_i \in C} Frequency_X(W_i)}{Size_X} \quad (1)$$

where $C = W_1, W_2, \dots, W_N$ represents a class of words.

The dominance score of the foreground corpus with respect to the background corpus is computed as a ratio between the coverages of the two corpora:

$$Dominance_F(C) = \frac{Coverage_F(C)}{Coverage_B(C)} \quad (2)$$

A dominance score equal or close to 1 for a given class indicates that the foreground and background corpora contain a similar distribution of words from the class.

A score greater than 1 indicates that the foreground corpus contains more words from a given class than the background corpus. In contrast, a value lower than 1 suggests that words from a class are underrepresented in the foreground corpus instead of the background corpus.

5. Results and Discussion

We computed dominance scores with each class in the foreground to reveal the dominant LIWC categories in the two corpora (depression and control). In Table 2 we have on the left side the analysis with the happy moments from the depression class as the foreground corpus and the happy moments of control subjects as the background corpus. On the right side of Table 2, the moments from the control users are the foreground corpus. From these results, it is apparent that depressed individuals more frequently manifest happy moments relating to family and friends (dominance scores > 1.4): *"I've been training regularly with my entire family for 3 months now and I love every bit of it."*, *"I'm with a woman I love and I have really supportive friends and family."*. Such events are more likely to be considered of greater value when compared to the happy moments of the control group. The control group, however, more frequently exhibit happy moments in everyday situations, in contexts related to entertainment, sports and financial plans: *"Knowing I make more money than you does make me happy."*, *"Eating pop corn and seeing movie."*, *"I bought a new controller and worked perfect."*. In line with psychology research, van Harmelen et al. (2016) found that the support of family and friends is negatively correlated with depression symptoms, in a study conducted on adolescents. Events related to positive experiences with loved ones are more noteworthy in a depressed subject's life, more so than in the case of control subjects.

Moreover, introspective discourse is more prevalent in depressed individuals (i.e., SELF category), which is in accordance with other works (Bucur et al., 2021b; Rude et al., 2004). Happy moments related to sleep are also more prevalent than those of the control category - poor sleep quality being one of the depression symptoms (O'Leary et al., 2017). Subjects diagnosed with depression also have more happy moments related to music (e.g., *"Together we crafted a legacy of music that has given so many people happiness and great memories."*, *"I read a bit from my book and listened to music."*), which is indicative of the positive effects music therapy has on mental well-being (Tang et al., 2020).

A surprising result was to find the death category from LIWC in the reasons for the happy moments of the control group. However, a closer manual analysis of the moments containing keywords from the death category revealed that, even if this category is comprised of death-related words (e.g., kill, dead, dying), they may not be used with their primary sense (e.g., *"I have been*

Depression			Control		
Category	Score	Example words	Category	Score	Example words
FRIENDS	1.81	friend, love, boyfriend, girlfriend, mate, bud, neighbor	TV	1.74	ad, show, movie, tv, actor, video
MUSIC	1.48	song, sing, band, rap, listen, music	BODY	1.21	eye, heal, breath, pregnant, cheek, skin
SLEEP	1.47	night, bed, sleep, dream, woke, nap	MONEY	1.41	fee, owe, rent, bought, pay, invest
FAMILY	1.45	dad, kin, ex, son, great, family, mom	DEATH	1.02	die, dead, decay, dying, ashes, terminate
SEXUAL	1.38	sex, bi, love, stud, fuck, hug	RELIGION	0.99	sin, christ, hell, god, angel, bless
SELF	1.36	I, me, my, we, us, our, mine	SPORTS	0.97	play, game, ski, team, running, sport, exercise

Table 2: Top classes and dominance scores for depression and control. Example words for each category are the ones found in our two corpora.

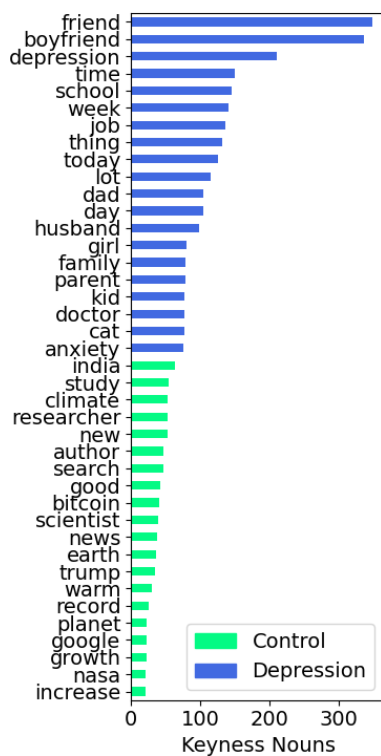


Figure 2: Keyness scores for nouns in the extracted happy moments from eRisk. Individuals with depression are more likely to discuss subjects related to family and friends than control users.

able to increase dead lifts from 70 kg to 100kg.”) or the words may be used in the context of video games (e.g., “I killed 5 King Black Dragons.”).

To further understand the differences in the happy moments reported by users with depression and control, we compute the keyness score (Kilgarriff, 2009; Gabrielatos, 2018) for the two corpora. We perform a keyness analysis by comparing the frequencies of words from the happy moments of users diagnosed

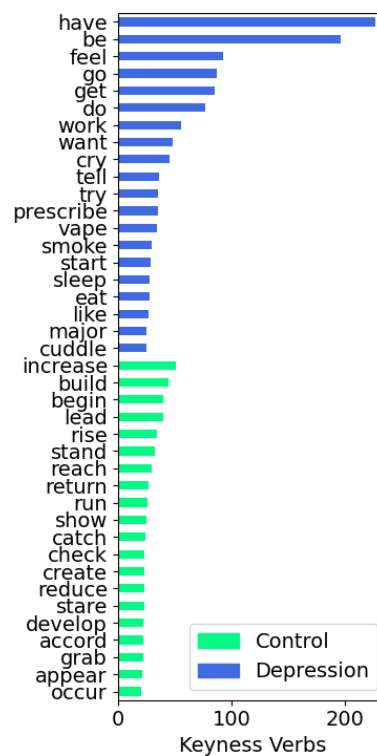


Figure 3: Keyness scores for verbs in the extracted happy moments from eRisk. Individuals with depression are more likely to use emotion-related verbs, while users from the control group use more action verbs.

with depression (target corpus) to the frequencies of words from the texts of control users (reference corpus). We perform this analysis on nouns and verbs. We present the top 20 nouns and verbs from the two corpora (depression and control) ordered by their log-likelihood ratio (G^2) (Dunning, 1993) in Figure 2 and 3.

In Figure 2 we show the most frequent nouns in the happy moments of individuals diagnosed with depres-

sion in comparison with the control group. A similar pattern of words related to family and friends is also found in the keyness analysis of words from the happy moments of users from the depression group.

In Figure 3 we present the most frequent 20 verbs from each group. Users diagnosed with depression use more verbs related to emotions (e.g., feel, cry, cuddle), as opposed to users from the control group, which use more action verbs (e.g., build, lead, run) in the texts of their happy moments.

6. Conclusion

In this work, we explored the reasons for happiness in social media text. Using the HappyDB dataset of happy moments and the Positive-Unlabeled Learning (PU-learning) framework, we automatically constructed a model to extract happy moments from the eRisk dataset. Our linguistic analysis through LIWC and keyness information showed that people diagnosed with depression more commonly experience joy from moments related to their family and friends and attribute greater significance to such events, while control subjects have happiness moments related to common everyday activities in contexts related to money, sports and entertainment.

This preliminary work is an exploration into manifestations of happiness of depressed individuals. The datasets used may be positively biased towards depressed individuals, which themselves attribute great value to their experiences of happiness, and choose to express it online; it is not clear the extent to which control users are expressing their happy moments online, rather than just enjoying them in the real world. However, our preliminary results are worthy of further development in future work, being in line with research performed in psychological studies (Tang et al., 2020; Rude et al., 2004), and could prove helpful in developing therapeutic chatbots and developing suitable positive psychological interventions. Furthermore, we believe that our results pave the way to a more in-depth analysis of expressions of happiness, by analyzing aspects of agency and sociality in happy moments. Moreover, due to computational limitations, we trained our PU-learning model on a sub-sample of the corpus; in future work, we aim to include the entire HappyDB dataset into the training data and perform our analyses on larger datasets (i.e. RSDD (Yates et al., 2017)).

7. Bibliographical References

Alarcon, G. M., Bowling, N. A., and Khazon, S. (2013). Great expectations: A meta-analytic examination of optimism and hope. *Personality and Individual Differences*, 54(7):821–827.

Aragon, M. E., Lopez-Monroy, A. P., Gonzalez-Gurrola, L.-C. G., and Montes, M. (2021). Detecting mental disorders in social media through emotional patterns—the case of anorexia and depression. *IEEE Transactions on Affective Computing*.

Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., Stepanov, D., Suhara, Y., Tan, W.-C., and Xu, Y. (2018). HappyDB: A corpus of 100,000 crowdsourced happy moments. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Bucur, A.-M. and Dinu, L. P. (2020). Detecting early onset of depression from social media text using learned confidence scores.

Bucur, A., Cosma, A., and Dinu, L. P. (2021a). Early risk detection of pathological gambling, self-harm and depression using BERT. *CoRR*, abs/2106.16175.

Bucur, A.-M., Podina, I. R., and Dinu, L. P. (2021b). A psychologically informed part-of-speech analysis of depression in social media. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 199–207, Held Online, September. INCOMA Ltd.

Bucur, A.-M., Zampieri, M., and Dinu, L. P. (2021c). An exploratory analysis of the relation between offensive language and mental health. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3600–3606, Online, August. Association for Computational Linguistics.

Burdisso, S. G., Errecalde, M., and Montes-y Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182–197.

Caragea, C., Dinu, L. P., and Dumitru, B. (2018). Exploring optimism and pessimism in twitter using deep learning. In Ellen Riloff, et al., editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 652–658. Association for Computational Linguistics.

Cobeli, S., Iordache, B., Yadav, S., Caragea, C., Dinu, L., and Iliescu, D. (2022). Detecting optimism in tweets using knowledge distillation and linguistic analysis of optimism. In *Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, France. European Language Resources Association (ELRA).

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

- Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1):71–75.
- Diener, E., Suh, E. M., Lucas, R. E., and Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological bulletin*, 125(2):276.
- Dunning, T. E. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1):61–74.
- D’raven, L. T. L., Moliver, N., and Thompson, D. (2015). Happiness intervention decreases pain and depression, boosts happiness among primary care patients. *Primary health care research & development*, 16(2):114–126.
- Elkan, C. and Noto, K. (2008). Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–220.
- Ford, B. Q., Shallcross, A. J., Mauss, I. B., Floerke, V. A., and Gruber, J. (2014). Desperately seeking happiness: Valuing happiness is associated with symptoms and diagnosis of depression. *Journal of Social and Clinical Psychology*, 33(10):890–905, December.
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In *Corpus Approaches to Discourse*, pages 225–258. Routledge.
- Hills, P. and Argyle, M. (2002). The oxford happiness questionnaire: a compact scale for the measurement of psychological well-being. *Personality and individual differences*, 33(7):1073–1082.
- Howes, C., Purver, M., and McCabe, R. (2014). Linguistic indicators of severity and progress in online text-based therapy for depression. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 7–16.
- Inkpen, D., Skaik, R., Gamaarachchige, P., Angelov, D., and Fredenburgh, M. T. (2021). uottawa at erisk 2021: Automatic filling of the beck’s depression inventory questionnaire using deep learning. *Working Notes of CLEF*, pages 21–24.
- Jaidka, K., Mumick, S., Chhaya, N., and Ungar, L. (2019). The cl-aff happiness shared task: Results and key insights. In *AffCon@ AAAI*.
- Kilgarriff, A. (2009). Simple maths for keywords. In *Proc. Corpus Linguistics*.
- Lara, J. S., Aragon, M. E., Gonzalez, F. A., and Montes-y Gomez, M. (2021). Deep bag-of-sub-emotions for depression detection in social media. *arXiv preprint arXiv:2103.01334*.
- Losada, D. E. and Crestani, F. (2016). A test collection for research on depression and language use. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 28–39. Springer.
- Mihalcea, R. and Liu, H. (2006). A corpus-based approach to finding happiness. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 139–144.
- Mihalcea, R. and Pulman, S. (2009). Linguistic ethnography: Identifying dominant word classes in text. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 594–602. Springer.
- Orabi, A. H., Buddhitha, P., Orabi, M. H., and Inkpen, D. (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*, pages 88–97.
- O’Leary, K., Bylsma, L. M., and Rottenberg, J. (2017). Why might poor sleep quality lead to depression? a role for emotion regulation. *Cognition and Emotion*, 31(8):1698–1706.
- Parks, A. C. and Layous, K. (2016). *Positive psychological interventions*. American Psychological Association.
- Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.
- Plath, S. and Hughes, T. (2000). *The unabridged journals of Sylvia Plath*. Anchor Books.
- Rude, S., Gortner, E.-M., and Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.
- Tang, Q., Huang, Z., Zhou, H., and Ye, P. (2020). Effects of music therapy on depression: A meta-analysis of randomized controlled trials. *PloS one*, 15(11):e0240862.
- Uban, A.-S. and Rosso, P. (2020). Deep learning architectures and strategies for early detection of self-harm and depression level prediction. In *CEUR Workshop Proceedings*, volume 2696, pages 1–12. Sun SITE Central Europe.
- van Harmelen, A.-L., Gibson, J. L., St Clair, M. C., Owens, M., Brodbeck, J., Dunn, V., Lewis, G., Croudace, T., Jones, P. B., Kievit, R. A., and Goodyer, I. M. (2016). Friendships and family support reduce subsequent depressive symptoms in at-risk adolescents. *PLOS ONE*, 11(5):1–20, 05.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063.
- Yates, A., Cohan, A., and Goharian, N. (2017). Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978, Copenhagen, Denmark, September. Association for Computational Linguistics.