Optimism, Pessimism, and the Language Between: Model Interpretability And Psycholinguistic Profiling

Ștefana-Arina Tăbușcă
Interdisciplinary School
of Doctoral Studies,
Faculty of Mathematics
and Computer Science,
University of Bucharest, Romania
stefana.tabusca@s.unibuc.ro

Liviu P. Dinu
Faculty of Mathematics
and Computer Science,
University of Bucharest, Romania
ldinu@fmi.unibuc.ro

Abstract

This study explores how optimism and pessimism are expressed in social media by combining psycholinguistic profiling with model interpretability. Using the OPT dataset, we fine-tune a RoBERTa-based classifier and apply LIME to examine both the most confident and the most ambiguous predictions. We analyze the influential tokens driving these decisions and identify lexical patterns linked to affective intensity, certainty, and social orientation. A complementary LIWC-based analysis of ground truth labels reveals systematic differences in emotional tone and cognitive style. PCA projections further show that optimism and pessimism occupy overlapping yet distinguishable regions in psycholinguistic space. Our findings demonstrate the value of linguistic interpretability in understanding dispositional sentiment.

1 Introduction

Optimism and pessimism - broadly defined as generalized expectations of positive or negative future outcomes - are well-established as psychologically meaningful traits (Scheier et al., 2001; Karhu et al., 2024). A large body of research links an optimistic outlook to better life outcomes, including longer life expectancy, improved physical health, and greater mental well-being (Tindle et al., 2012; Scheier et al., 1986). On the other hand, dispositional pessimism is associated with negative affect and vulnerability to depression; it can exacerbate stress, induce suicidal ideation, and adversely affect not only individuals but also their families and social circles (Zenger et al., 2010; Herwig et al., 2009; Hobbs et al., 2022). Maintaining an optimistic attitude has been shown to reduce stress and improve overall resilience (Malouff and Schutte, 2016; Shapira and Mongrain, 2010). These findings show that optimism and pessimism are not just internal dispositions - they are socially consequential and deeply tied to mental health (Broome et al., 2015; Technow et al., 2015; Kronström et al., 2011).

Social media platforms offer unprecedented access to public self-expression, providing a rich substrate for analyzing dispositional signals through language. Prior studies have leveraged social media text to predict personality traits, depressive symptoms, and subjective well-being (Zimmermann et al., 2016; Bucur et al., 2021, 2022; Xu et al., 2024). Detecting optimism and pessimism in online discourse thus holds practical value for understanding and supporting psychological health at scale, such as identifying at-risk individuals, monitoring community sentiment, or amplifying positive discourse (Gan et al., 2024; Zhang et al., 2022).

This paper addresses the task of identifying optimism and pessimism in social media, using the Optimism-Pessimism Twitter (OPT) dataset introduced by Ruan et al. (2016), which contains tweets annotated for degrees of optimistic or pessimistic expression. We use this dataset to explore how optimism and pessimism are expressed linguistically, how they can be modeled using transformer-based architectures, and how such models can be interpreted using psycholinguistic and local explanation tools.

The study is guided by the following research questions:

1. (RQ1) What psycholinguistic features distinguish optimistic from pessimistic language in social media?

2. (RQ2) To what extent can local interpretability methods reveal the linguistic indicators that models rely on for their most confident predictions, as well as for their most ambiguous ones?

Proceedings of Recent Advances in Natural Language Processing, pages 1224–1231
Varna, Sep 8–10, 2025
https://doi.org/10.26615/978-954-452-098-4-141

By addressing these questions, we aim to deepen the understanding of how outlook is encoded in language, rather than solely improving classification performance. The combination of classification, interpretation, and psycholinguistic profiling offers a comprehensive lens on the language of optimism and pessimism in social media contexts.

2 Related Work

Early computational studies approached optimism–pessimism classification with traditional machine learning techniques. Ruan et al. (2016) were among the first to tackle this problem on social media, creating the OPT Twitter corpus and training classifiers to distinguish optimistic vs. pessimistic users. In their study, each user's tweets were aggregated and annotated on a scale from very pessimistic (-3) to very optimistic (+3). They experimented with standard algorithms like Naïve Bayes, decision trees, and SVMs using bag-of-words features.

Caragea et al. (2018) used neural network models to analyze tweets to identify optimism and pessimism. They employed bidirectional LSTMs (Long Short-Term Memories), CNNs (Convolutional Neural Networks), and stacked GRUs (Gated Recurrent Units) to outperform traditional baselines. They also investigated whether optimism/pessimism could be detected through generic sentiment analysis and found that this task is not reducible to positive vs. negative sentiment alone. The study further analyzed verb tense usage and polarity words in optimistic and pessimistic tweets, revealing a higher prevalence of present-tense verbs in optimistic posts and slightly more past-tense verbs in pessimistic ones. These findings suggest the importance of psycholinguistic features in characterizing the language of optimism.

Transformer language models have improved the accuracy of optimism detection by applying large pre-trained models like BERT and XLNet to the OPT benchmark. Alshahrani et al. (2020) fine-tuned XLNet for optimism/pessimism classification and introduced an ensemble "deep consensus" strategy, achieving substantial gains over earlier deep neural networks. In a follow-up, Alshahrani et al. (2021) employed BERT and a soft label assignment technique to account for annotation uncertainty, which further boosted performance. These results highlight the benefits of transformer encoders in capturing nuanced linguistic cues of optimism, even with small datasets.

Beyond single-task models, researchers have explored multi-task and cross-domain approaches. One recent study by Cobeli et al. (2022) proposed a multi-task knowledge distillation (MTKD) framework to leverage related tasks such as sentiment analysis and hate speech detection for improving optimism detection. By transferring knowledge from auxiliary tasks, their student model gained a richer understanding of language cues, achieving significant improvements.

Blanco and Lourenço (2022) used a deep learning model to analyze Twitter conversations about COVID-19, revealing how public optimism fluctuated during pandemic waves. They used lexicon-based sentiment signals and an interpretable model to explain changes, demonstrating the social insights these models can provide.

The push for explainable AI has led to the integration of NLP (Natural Language Processing) tools like LIME (Ribeiro et al., 2016) into tweet classification pipelines. This allows researchers to uncover the "black box" of words driving a model's prediction, increasing trust in sentiment analysis. LIME has been used in optimism/pessimism detection by Blanco and Lourenço (2022), identifying salient terms that signal optimism or pessimism in specific examples. These local explanations are crucial for dealing with psychological traits.

3 Data

The preeminent dataset for optimism/pessimism identification currently, introduced by Ruan et al. (2016), is annotated from Twitter and employed for the optimism/pessimism detection task, comprising 7,475 randomly selected tweets from 500 pessimistic individuals and 500 deemed optimists. Texts were selected by identifying tweets that had terms connected to optimism or pessimism, therefore emphasizing both optimistic and pessimistic people. The aggregation of their recent tweets yielded a comprehensive dataset. Using Amazon Mechanical Turk, human annotators evaluated and classified each tweet according to a disposition scale ranging from 3 (extremely optimistic) to -3 (very pessimistic). This scale allowed for complex differentiation across tweets and the recognition of different degrees of optimism and pessimism in the text. The average of all the evaluations for the acquired annotations is the final score. Accuracy was ensured using quality control procedures such as correctly defining optimism and pessimism, excluding raters who gave incorrect responses to established "check" questions, and comparing annotations to the average score to spot anomalies.

In our research, we investigate the binary class interpretation, classifying tweets as optimistic if their scores are greater than 0 and pessimistic if their scores are 0 or below. This setup yielded the ratio of 4187 optimistic and 3288 pessimistic posts.

4 Methodology

4.1 Model Development

To tackle the classification task, we fine-tuned a RoBERTa-based architecture using the HuggingFace Transformers library, initializing from the twitter-roberta-base-sentiment-latest checkpoint (Camacho-Collados et al., 2022). We chose RoBERTa as the base model due to its strong performance on a wide range of NLP classification tasks, particularly sentiment analysis, which aligns closely with optimism/pessimism detection. Additionally, this particular model is pretrained on approximately 124 million tweets collected between January 2018 and December 2021 and further optimized on the TweetEval sentiment analysis benchmark (Barbieri et al., 2020), making it well-suited for social media text, especially Twitter/X as in our setup.

For fine-tuning, we adopted a standard configuration, as per the HuggingFace setup: a learning rate of 5e-5, batch size of 8, and a maximum sequence length of 128 tokens. Training was conducted over three epochs with the AdamW optimizer, which integrates weight decay to regularize training. We applied a linear learning rate scheduler with warm-up, reserving 10% of the steps for warm-up before linearly decreasing the rate to zero. To improve stability, the gradient norm was clipped at a maximum of 1.0. A fixed random seed (42) was used to ensure reproducibility. Early stopping was enabled to prevent overfitting, with training halted if no improvement was observed after five consecutive evaluations. An early stopping threshold of 0.01 was used to define the minimum required change in the monitored validation metric.

4.2 Explainability with LIME

To gain deeper insight into the linguistic signals driving model predictions, we employed LIME (Local Interpretable Model-Agnostic Explanations) to extract token-level contributions to classification decisions. LIME approximates the local decision boundary of the model for individual instances by perturbing the input and learning a sparse, interpretable linear model that highlights which features (in our case, tokens) most influence the prediction. This allowed us to identify which specific words drove the model's decision toward either the optimistic or pessimistic class.

We focused our interpretability analysis along two axes: (1) confidently predicted examples (i.e., those with prediction probabilities far from 0.5), and (2) ambiguous or borderline cases, including those with near-random confidence scores or misclassified labels. This dual perspective allowed us to isolate the prototypical lexical drivers of sentiment on one hand, and on the other, examine the words responsible for confusion or conflict in model behavior.

4.3 LIWC Data Profiling

To investigate how optimistic and pessimistic language differs along psychological and emotional lines, we employed the LIWC-22 (Linguistic Inquiry and Word Count) framework (Boyd et al., 2022) to extract a comprehensive set of psycholinguistic features from all tweets in the dataset. The present analysis focuses exclusively on the distribution of linguistic features with respect to the round truth labels in the binary setup.

The LIWC framework provides over 90 validated dimensions, including affective categories (e.g., positive emotion, anxiety, sadness), cognitive mechanisms (e.g., causation, certainty, tentativeness), and grammatical categories (e.g., pronouns, verbs, negations). We excluded non-informative metadata and structural columns, retaining only those features that reflect linguistic content.

We computed the mean value of each LIWC dimension separately for optimistic and pessimistic tweets. To visualize the features contributing most strongly to the distinction between these two sentiment categories, we calculated the absolute difference in mean scores across the two groups and ranked them. The twenty most divergent features were selected for comparative analysis.

Additionally, we applied Principal Component Analysis (PCA) to reduce the dimensionality of the LIWC space and facilitate a visual exploration of tweet distribution. Each tweet was represented as a vector in the LIWC feature space, and PCA was used to project this space into two dimensions while retaining as much variance as possible. The resulting scatterplot displays how tweets cluster or disperse based on their psycholinguistic attributes, with colors indicating their sentiment label.

5 Results

5.1 Model Performance

As previously stated, this study takes into account the binary 0-threshold scenario based on how the OPT dataset's annotation scores are interpreted. The data was split 70%/15%/15% for train/validation/test respectively. The performance of our deployed model (which we dub RoBERTa-OPT-BinaryLabels) during the training phase, the test outcomes, and the baselines will be discussed next. Metrics including Loss, F1 Score, Precision, Recall, and Accuracy were used to assess the model. The results are displayed in Table 1:

Metrics	Validation	Test
Loss	0.4674	0.4834
F1 Score	0.8415	0.8492
Precision	0.8603	0.8828
Recall	0.8234	0.8180
Accuracy	0.8323	0.8307

Table 1: Results for the binary setting model, RoBERTa-OPT-BinaryLabels

The RoBERTa-OPT-BinaryLabels model achieved high accuracy and precision, with low loss values indicating stable training. The model's optimistic class achieved high confidence in positive predictions, with recall values above 81.5%, indicating a strong ability to identify optimistic tweets while minimizing false negatives. Although performance improvement was not the primary goal, these metrics confirm the model's well-calibration and competitiveness. The model performs comparably to recent state-of-the-art systems in tweet-level binary classification, setting a strong baseline for further interpretability and linguistic analysis.

Model	Val. Acc.
Naïve Bayes Caragea et al. (2018)	84.10
SVM Caragea et al. (2018)	83.30
CNN Caragea et al. (2018)	90.32
BiLSTMN Caragea et al. (2018)	87.24
GRUStack Caragea et al. (2018)	87.76
XLNet-Base Alshahrani et al. (2020)	96.16
XLNet-Large Alshahrani et al. (2020)	96.45
BERT-Base with SLA Alshahrani et al. (2021)	-
MTKD OPT + Hate +Sent Cobeli et al. (2022)	85.14
RoBERTa-OPT-BinaryLabels	83.22

Table 2: Baselines and Validation Accuracy comparison with RoBERTa-OPT-BinaryLabels

Model	Test Acc.
Naïve Bayes Caragea et al. (2018)	74.20
SVM Caragea et al. (2018)	67.80
CNN Caragea et al. (2018)	77.78
BiLSTMN Caragea et al. (2018)	79.65
GRUStack Caragea et al. (2018)	80.19
XLNet-Base Alshahrani et al. (2020)	84.25
XLNet-Large Alshahrani et al. (2020)	85.28
BERT-Base with SLA Alshahrani et al. (2021)	85.69
MTKD OPT + Hate +Sent Cobeli et al. (2022)	86.60
RoBERTa-OPT-BinaryLabels	83.06

Table 3: Baselines and Test Accuracy comparison with RoBERTa-OPT-BinaryLabels

5.2 LIME Analysis

In confidently predicted posts, LIME consistently highlighted high-valence sentiment words that aligned well with the model's final decision (Figure 1. Optimistic instances were characterized by tokens such as "amazing," "excited," "happy," "legend," "love," and "awesome"-terms strongly associated with enthusiasm, admiration, and joy. Notably, the presence of proper nouns like "rihanna" and "avatar" suggests that cultural references and celebrity names can function as emotionally charged cues, acting as indirect proxies for affective content.

In contrast, pessimistic predictions were most strongly influenced by emotionally intense and often profane language. Words like "fuckboy," "shitting," "useless," "hate," and "kill" dominated the explanations for these examples. Many of these carry not only negative valence but also moral judgment or aggression, indicating that pessimism in this dataset may be linguistically tied to expressions of frustration, disappointment, or existential despair. The token "never", which is syntactically neutral but semantically absolute, also appeared with significant weight, suggesting that linguistic absolutes may correlate with negative sentiment in user-generated text.

These results suggest that the model has learned to rely on a core lexicon of affectively polar tokens. The interpretability afforded by LIME makes these learned associations transparent, offering external validation for the semantic features learned during training.

We also analyzed posts that were predicted with low confidence or misclassified, defined as those whose predicted probability scores fell near the midpoint threshold (e.g., within the 0.48–0.52 range). These cases often included tokens with high individual weights pulling in opposite directions, reflecting genuine semantic ambiguity or con-

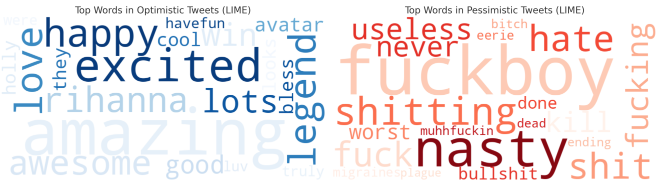

Figure 1: Word clouds of the most predictive tokens for optimism (left) and pessimism (right) based on LIME explanations. Token sizes reflect the average LIME weight assigned by the model to each word in its most confidently predicted posts.

Figure 2: Word cloud of the most influential tokens in ambiguous tweets, as identified by LIME. Word size reflects the absolute magnitude of each token's contribution to model predictions.

textual contradiction (most predictive tokens visible in Figure 2). For example, in one ambiguous post the words "sun" and "amazing" were weighted positively, while "sad" and "can't" exerted a strong negative influence. Such examples illustrate a form of internal polarity conflict-where the lexical signals point simultaneously toward optimism and pessimism-likely contributing to prediction uncertainty.

In these borderline cases, influential tokens were not neutral or semantically irrelevant (as might be assumed) but rather emotionally suggestive in opposing directions. In this light, misclassifications may reflect genuine polysemy, sarcasm, or mixed sentiment expressions.

5.3 LIWC Results and Clustering

The analysis reveals distinct patterns in how optimistic and pessimistic sentiment is expressed at a linguistic level, with the distribution of the top features for both classes visible in Figure 3.

Optimistic tweets tend to contain higher proportions of words associated with positive emotion, certainty, and affiliation, suggesting a focus on connection, confidence, and constructive framing. These tweets also show elevated scores in social and communal language, including references to friends and inclusivity.

In contrast, pessimistic tweets exhibit greater use of terms related to negative emotion, particularly those associated with sadness, anxiety, and anger. They also show a stronger presence of cognitive process features, such as causation and insight, which may reflect rumination or evaluative thought patterns common in expressions of frustration or resignation.

The PCA projection (visible in Figure 4) shows that distinct clustering does occur, however, the two sentiment categories are not fully separable in LIWC space. Many tweets appear in a shared region, indicating a degree of lexical and psychological overlap. This is consistent with the nature of informal text on platforms like Twitter, where sentiment may be mixed, ironic, or contextually dependent.

These findings suggest that although psycholinguistic profiles provide valuable signals for distinguishing optimism and pessimism, the boundary between them is not always sharply defined. Emotional language is frequently layered and dynamic, which has direct implications for the design of sentiment classification models and the interpretation of their outputs.

6 Conclusions

This study offers three key takeaways for sentiment modeling. First, token-level explanations from LIME provide interpretable insights into the model's decision-making and help verify alignment with human intuition. Second, analyzing ambiguous predictions reveals where the model struggles, emphasizing the value of interpretability for understanding uncertainty and error. Third, the model's sensitivity to specific word classes-such as affective intensifiers, absolutes, and culturally loaded terms-suggests concrete linguistic cues that could

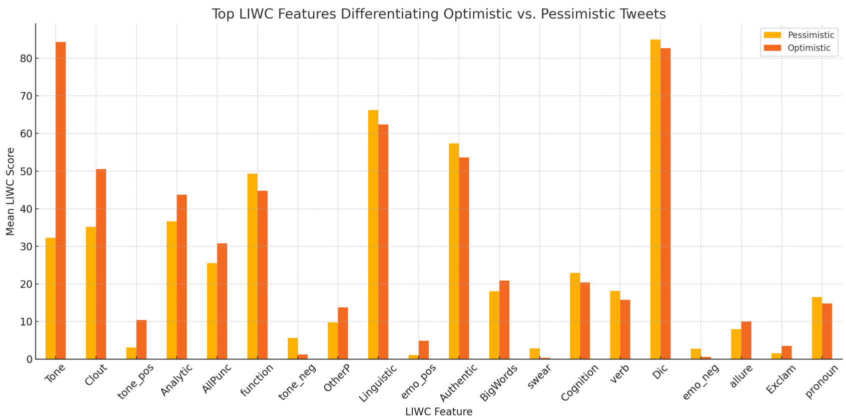

Figure 3: Top LIWC features that differentiate optimistic and pessimistic tweets, based on absolute mean difference.

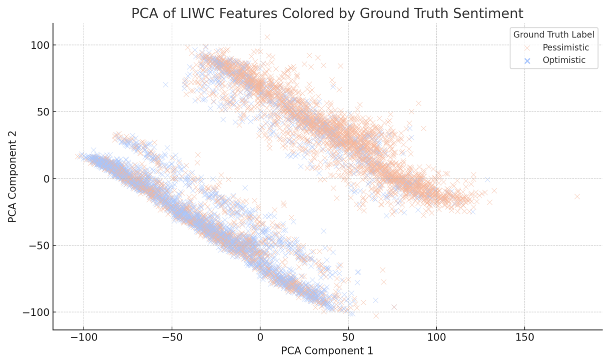

Figure 4: PCA projection of tweets in LIWC feature space, colored by optimism and pessimism labels.

inform future hybrid or rule-based systems.

Beyond model-level interpretation, the LIWC-based analysis shows that optimism and pessimism are linguistically and cognitively distinct. Optimistic language is marked by positive affect, certainty, and affiliation, while pessimistic posts exhibit more negative emotion, cognitive complexity, and tentativeness. PCA projections of LIWC features further reveal that these outlooks, while overlapping, form distinguishable clusters in psycholinguistic space.

Together, these findings show that interpretability and linguistic profiling offer meaningful insights into both model behavior and the structure of dispositional sentiment in language.

Limitations

While this study is centered on the OPT dataset, its focused design enables a controlled and linguistically rich exploration of optimism and pessimism in a well-defined social media context. The analysis is limited to English tweets and post-level features, which ensures clarity and comparability but may not capture cross-lingual or longitudinal dynamics. LIME, though a strong interpretability tool, can introduce minor variability in explanations; however, it remains effective for identifying salient lexical patterns in both confident and ambiguous cases. These design choices reflect a deliberate emphasis on linguistic interpretability over generalization.

Ethics Statement

We analyze publicly available tweets without collecting personal or identifiable user data. The study is strictly observational and not intended for diagnostic use. Any application of such models to mental health contexts must ensure transparency, human oversight, and careful mitigation of risks like misclassification or stigmatization.

Acknowledgments

This work was supported by the project "Romanian Hub for Artificial Intelligence - HRIA”, Smart Growth, Digitization, and Financial Instruments Program, 2021-2027, MySMIS no. 334906.

References

A. Alshahrani, M. Ghaffari, K. Amirizirtol, and X. Liu. 2020. Identifying optimism and pessimism in twitter messages using xlnet and deep consensus. In *International Joint Conference on Neural Networks (IJCNN)*.

A. Alshahrani, M. Ghaffari, K. Amirizirtol, and X. Liu. 2021. Optimism/pessimism prediction of twitter messages and users using bert with soft label assignment. In *International Joint Conference on Neural Networks (IJCNN)*.

Fabio Barbieri, José Camacho-Collados, Lisa Ellen Anke, and Leandro Neves. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics.

G. Blanco and A. Lourenço. 2022. Optimism and pessimism analysis using deep learning on covid-19 related twitter conversations. *Information Processing and Management*, 59(3):102918.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10.

M. R. Broome, K. E. A. Saunders, P. J. Harrison, and S. Marwaha. 2015. Mood instability: Significance, definition and measurement. *British Journal of Psychiatry*, 207(4):283–285.

A. Bucur, A. Cosma, and L. P. Dinu. 2022. Life is not always depressing: Exploring the happy moments of people diagnosed with depression. *arXiv (Cornell University)*.

A. Bucur, I. R. Podina, and L. P. Dinu. 2021. A psychologically informed part-of speech analysis of depression in social media. In *International Conference on Recent Advances in Natural Language Processing*, pages 199–207.

J. Camacho-Collados, K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. E. Anke, F. Liu, and E. M. Camara. 2022. Tweetnlp: Cutting-edge natural language processing for social media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

C. Caragea, L. P. Dinu, and B. Dumitru. 2018. Exploring optimism and pessimism in twitter using deep learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ş. Cobeli, I. Iordache, S. Yadav, C. Caragea, L.P. Dinu, and D. Iliescu. 2022. Detecting optimism in tweets using knowledge distillation and linguistic analysis of optimism. In *International Conference on Language Resources and Evaluation*.

L. Gan, Y. Guo, and T. Yang. 2024. Machine learning for depression detection on web and social media. *International Journal on Semantic Web and Information Systems*, 20(1):1–28.

U. Herwig, A. B. Brühl, T. Kaffenberger, T. Baumgärtner, H. Boeker, and L. Jäncke. 2009. Neural correlates of 'pessimistic' attitude in depression. *Psychological Medicine*, 40(5):789–800.

Catherine Hobbs, Petra Vozarova, Aarushi Sabharwal, Punit Shah, and Katherine Button. 2022. Is depression associated with reduced optimistic belief updating? *Royal Society Open Science*, 9(2):190814.

Jutta Karhu, Juha Veijola, and Mirka Hintsanen. 2024. The bidirectional relationships of optimism and pessimism with depressive symptoms in adulthood–a 15-year follow-up study from northern finland birth cohorts. *Journal of Affective Disorders*, 362:468–476.

Kim Kronström, Hasse Karlsson, Hermann Nabi, Tuula Oksanen, Paula Salo, Noora Sjösten, Marianna Virtanen, Jaana Pentti, Mika Kivimäki, and Jussi Vahtera. 2011. Optimism and pessimism as predictors of work disability with a diagnosis of depression: a prospective cohort study of onset and recovery. *Journal of affective disorders*, 130(1-2):294–299.

J. M. Malouff and N. S. Schutte. 2016. Can psychological interventions increase optimism? a meta-analysis. *The Journal of Positive Psychology*, 12(6):594–604.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM.

X. Ruan, S. R. Wilson, and R. Mihalcea. 2016. Finding optimists and pessimists on twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.

M. F. Scheier, J. K. Weintraub, and C. S. Carver. 1986. Coping with stress: Divergent strategies of optimists and pessimists. *Journal of Personality and Social Psychology*, 51(6):1257–1264.

Michael F Scheier, Charles S Carver, and Michael W Bridges. 2001. Optimism, pessimism, and psychological well-being. *Optimism & pessimism: Implications for theory, research, and practice.*, pages 189–216.

L. B. Shapira and M. Mongrain. 2010. The benefits of self-compassion and optimism exercises for individuals vulnerable to depression. *The Journal of Positive Psychology*, 5(5):377–389.

J. R. Technow, N. A. Hazel, J. R. Abela, and B. L. Hankin. 2015. Stress sensitivity interacts with depression history to predict depressive symptoms among youth: prospective changes following first depression onset. *J Abnorm Child Psychol.*, 43(3):489–501.

Hilary Tindle, Bea Herbeck Belnap, Patricia R Houck, Sati Mazumdar, Michael F Scheier, Karen A Matthews, Fanyin He, and Bruce L Rollman. 2012. Optimism, response to treatment of depression, and rehospitalization after coronary artery bypass graft surgery. *Psychosomatic medicine*, 74(2):200–207.

Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K. Dey, and Dakuo Wang. 2024. Mental-llm: Leveraging large language models for mental health prediction via online text data. 8(1).

M. Zenger, H. Glaesmer, M. Höckel, and A. Hinz. 2010. Pessimism predicts anxiety, depression, and quality of life in female cancer patients. *Japanese Journal of Clinical Oncology*, 41(1):87–94.

T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou. 2022. Natural language processing applied to mental illness detection: A narrative review. *npj Digital Medicine*, 5(1).

J. Zimmermann, T. Brockmeyer, M. Hunn, H. Schauenburg, and M. Wolf. 2016. First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical Psychology Psychotherapy/Clinical Psychology and Psychotherapy*, 24(2):384–391.