

Automated detection of developer mental health status in collaboration platforms using NLP and activity analysis

Abewardhana M.R.S.C.^{1*}, Ishanka U.A.P.²

¹Department of Software Engineering, Faculty of Computing,
Sabaragamuwa University of Sri Lanka, Sri Lanka

²Department of Data Science, Faculty of Computing,
Sabaragamuwa University of Sri Lanka, Sri Lanka

*sahanchinthaka04@gmail.com

Software developers experience significant mental health challenges due to work-related stressors, making it essential for workplaces to adopt systematic and proactive strategies for identifying and supporting employees with psychological difficulties. This study proposes a novel NLP-based model that predicts developers' mental health by analyzing the emotional patterns embedded in their chat messages. The methodology consists of two main stages: (1) categorizing developers' messages into emotion classes using fine-tuned language models and (2) evaluating monthly emotion distributions to derive individualized mental health risk scores. A fine-tuned DistilRoBERTa-base model, trained and evaluated on the Gitter Emotion Dataset, achieved a weighted accuracy of 62.8% on developer-specific messages, demonstrating the feasibility of domain-adapted emotion classification for software engineering communication. Using 30-day holistic assessments, mental health scores ranging from 38 to 57 were produced, effectively differentiating developers exhibiting a higher proportion of negative emotional expressions from those whose communication patterns were predominantly neutral or positive. This research provides a continuous, non-intrusive method for collecting and interpreting mental health signals by leveraging communication tools that software developers already use as part of their daily workflow. This study builds upon previous studies in this area, all of which have used either facial emotion recognition or text analysis derived from social media to do their work, by being the first study to use developer-centric chat to connect emotion expressed in writing to subsequent mental health assessments. As demonstrated in this study, emotion expressed in text in the context of software development can be reliably used as an indirect measure of an individual's mental health. Therefore, the scope of emotion-based mental health rating methodologies can be extended beyond simply using expressive face information. The model developed in this study has implications for both practitioners wanting to monitor employees' well-being at work, as well as researchers in the field of computational psychiatry by demonstrating that the emotion represented in domain specific text can serve as a proxy for assessing severity of an individual's mental health.

Keywords: *Developer mental health, Text emotion classification, Natural language processing, Occupational wellness, Predictive modeling*