Preregistration Protocol: Smartphones as mood barometers: Predicting mood in daily life using different sensing modalities

This preregistration protocol deals with specific research questions and has been completed after the data collection. Test data were available to specify the preregistration. Study procedures and further background information are described in the corresponding basic protocol. This template is inspired by the OSF Prereg Challenge template (https://osf.io/).

Working Title
Smartphones as mood barometers: Predicting mood in daily life using different sensing modalities

Author(s) of the preregistration protocol
Fiona Kunz

Date
31.05.2022

Abstract/ Summary of Study Description
Momentary experiences of positive and negative emotionality are core components of well-being and performance. This study investigates whether passively sensed smartphone data can be used to recognize individuals' mood (i.e. Valence and Arousal (Russell, 1980)) based on their smartphone sensing data. The exploratory analysis uses data generated from *N* = 453 participants in a two-week experience sampling wave which was part of the Smartphone Sensing Panel Study (SSPS; Schödel & Oldemeier, 2020). Different cross-validated machine learning algorithms are compared to predict participants' current mood given a variety of situational and behavioral variables, reflected by different smartphone

sensing modalities. Moreover, the impact of different time perspectives (i.e. daily versus hourly) on the predictive performance is investigated.

Keywords: *Smartphone sensing; mood; machine learning; predictive modeling.*

Background

Background Information (Optional; Short description of the theoretical background/introduction to research question)

Theoretical Background

From a psychobiological perspective, positive mood serves as an important proxy of mental well-being, ranging from the ability to withstand daily stress (e.g. Ong et al., 2006) to mental disorders like depression or anxiety (e.g. Kashdan & Steger, 2006; Wichers et al., 2020). Mental well-being in turn has positive impacts on objective health outcomes fostered by an improved immune system functioning and reduced risk of adverse physical health outcomes (e.g. Aichele et al., 2016; Steptoe et al., 2009; Veenhoven, 2008; see Howell et al. (2007) for a comprehensive overview).

Given its importance, prior research has therefore intensively investigated physiological markers of mood, focusing on EEG (e.g. Gable et al., 2021; Petrantonakis & Hadjileontiadis, 2010; Stikic et al., 2014), skin conductance or temperature (e.g. Sano et al., 2015; 2018; Steptoe et al., 2005; 2009). However, sensing these physiological markers typically requires installing, wearing, or otherwise carrying external sensors and devices, which limits their applicability to everyday life. The pervasiveness of modern technologies like smartphones has enabled the timely delivery of experience samplings, allowing the real-time collection of (self-reported) psychological outcomes combined with passively collected sensor-based data. This has led to the development of several mobile phone applications that prompt their users to assess and report their mood one or more times per day, using one or more different scales (e.g. *DeepMood* app, see Cao et al. (2017); *Emotion Sense* app, see Servia-Rodríguez et al., (2017); *MoodExplorer* app, see Zhang et al. (2018); *MoodScope* app, see LiKamWa et al. (2013)).

Trying to “put mood into context” (Sandstrom et al., 2017), research has shown that passively, timely and accurately sensed data of different types can serve as indicators for an individual’s momentary mood. For example, GPS and accelerometer data, including places visited (Boukhechba et al., 2017; Chow et al., 2017; Müller et al., 2020; Sandstrom et al., 2017), as well as mobility patterns (Cai et al., 2018; DeMasi et al., 2017; Lee et al., 2017; Ren et al., 2022; Spathis et al., 2019) were shown to be useful in recognizing mood in daily life. Moreover, previous research has linked communication behavior (e.g. based on text messages or call data) to mood (Boukhechba et al., 2017; Ma et al., 2012; Sano et al., 2018; Servia-Rodríguez, et al., 2017). Other studies for example focused on sensing the user's mood based on smartphone usage patterns like screen time or app usage (e.g. Cao et al., 2017; LiKamWa et al., 2013; Messner et al., 2019; Ren et al., 2022; Sano et al., 2018). An extensive overview of features used in previous wellbeing and affect studies is shown in the table 1 in the additionally uploaded document "*MOOD_preregistration_features.pdf*".

In addition to the behavioral and situational factors described above, person-related factors such as personality traits (e.g. emotional stability or neuroticism) have been strongly associated with self-reported mood (e.g. Cheng & Furnham, 2003; Ching et al., 2014; Geukes et al., 2017). Concretely, previous studies have provided preliminary evidence of increased predictive accuracy when personality self-reports are additionally included in models predicting affect (e.g. Denissen et al., 2008; Kööts et al., 2011; Sandstrom et al., 2017).

Thus, while research has shown that smartphones can be harnessed as instruments for unobtrusive monitoring of mood, the present study contributes to this research in the following ways.

First, the study compares different time perspectives of Arousal and Valence experience. Building on the widespread conceptualization of mood as relatively enduring wellbeing experience or mental health status (e.g. Canzian et al., 2015; Spathis et al., 2019; Wang et al., 2014; 2016), this study additionally focuses on the prediction of momentary mood experience (e.g., Wednesday at 5:15 pm) by using contextual information from passive sensing. Accordingly, this research does not only focus on daily behavioral or situational patterns like the number of locations visited per day (e.g. Ma et al., 2012, Servia-Rodríguez, et al., 2017), but additionally "zoom" into the participant's smartphone sensing data one hour before the respective experience sampling event.

Second, a variety of different sensing modalities is considered, reflecting different domains of behavioral and situational patterns (e.g. communication, mobility, music consumption, smartphone usage, weather, …) to uncover the full potential of smartphone sensing data for tracking an individual's experience of mood in daily life.

Research question(s)
Concretely, the following research questions are explored: (1) Can people's self-reported mood experience in daily life be predicted from smartphone sensing data? a. Can people's **momentary mood** be predicted from smartphone sensing data logged within a one-hour time window before the mood self-report? b. Can people's **daily mood** (i.e. daily average of momentary experience of emotional Valence and Arousal) be predicted from smartphone sensing data logged during that day? (2) Does the predictive performance improve when **personality traits** are additionally included as predictor variables in the model? (3) How are **different sensing modalities** (i.e. domains of behavioral and situational patterns) related to the predictability of (a) **momentary** and (b) **daily** mood (i.e. Valence and Arousal)?

Hypotheses *Please provide hypothesis for predicted results. If multiple hypotheses, uniquely number them (e.g. H1, H2a, H2b,) and refer to them the same way at other points in the registration document and in the manuscript.*
This study is exploratory in nature and uses an exploratory machine learning approach to investigate the research questions. The smartphone-sensed variables[1] used in the predictive modeling are derived from previous empirical studies (see *Appendix Table 1*). The procedure is preregistered as a transparent account of the research work.

[1] As it is common in the machine learning context, the term features or predictors will be used instead of variables in the following.

Variables

Which variables will be used? (see Variables in the basic protocol for an extensive overview of all available variables) *This section shall be used to unambiguously clarify which variables are used to operationalize the specified hypotheses. Please (a) list all variables that will be used in this study and (b) explicitly state the functional role of each variable (i.e., independent variable, dependent variable, covariate, mediator, moderator). It is important to (c) specify for each hypothesis how it is operationalized, i.e., which variables will be used to test the respective hypothesis and how the hypothesis will be operationally defined in terms of these variables. This section is closely related to the statistical models used to test the hypotheses.*

Data collection occurred as part of a the siy-month Smartphone Sensing Panel Study (SSPS) conducted by researchers at Ludwig-Maximilians-Universität München (LMU) in cooperation with Leibniz-Institut für Psychologie (ZPID) from May until November 2020 (for more details see Schödel & Oldemeier, 2020). The study was approved by the ethics board of the LMU Munich. The SSPS included three data collection modalities: (1) smartphone sensing, (2) experience sampling, and (3) monthly online surveys. Drawing from this extensive dataset, this study will analyze experience sampled self-reports as well as passively sensed data from smartphones owned by a representative sample of N = 453 participants collected in Germany.

Self-report Measures

Self-report measures analyzed in this study were collected in the second experience sampling wave (21.09.2020 to 04.10.2020) as well as survey one (18.05.2020 to 25.05.2020) and survey four (17.08.2020 to 23.08.2020) of the SSPS (see Schödel & Oldemeier, 2020).

In a two-week experience sampling phase, participants assessed their current mood between two to four times per day on the two dimensions Valence and Arousal. Concretely, participants were asked to indicate their Valence (*"How do you assess your current emotional state?"*) and Arousal (*"How do you assess your current activity level"*) level on a six-point Likert scale (0 = *very unpleasant / very inactive*, 1 = *unpleasant / inactive*, 2 = *rather unpleasant / rather inactive*, 3 = *rather pleasant / activated*, 4 = *pleasant / activated*, 5 = *very pleasant / very activated*). The two-dimensional assessment of the current state affect is oriented on the circumplex model of emotion developed by Russell (1980) and was chosen to keep participants' burden on an acceptable level, as items were presented among several other self-report measures (Schödel & Oldemeier, 2020).

In survey one of the SSPS, participants' *demographic characteristics* (e.g., age, gender, education level, nationality, marital status, employment status, etc.) were assessed. In survey four, participants' *Big-Five personality* (Open-Mindedness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) was measured via the German version of the Big Five Inventory, extra-short form (BFI-2-XS; Rammstedt et al., 2020). A total of 15 items measure the Big Five traits with three facets each which are rated on a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree.*

Smartphone-sensed Measures

Participants downloaded the PhoneStudy research app (an Android logging app for Android OS version 5 or higher) which collected data from the participant's phone's sensors and logs (for a comprehensive list of all logging event specifications, see Schödel & Oldemeier, 2020). This study uses a broad range of smartphone sensed behavioral and situational variables (e.g., calls, texts, app usage, connected devices, visited locations, etc.). All smartphone sensed data are timestamped as they were logged event-based (by occurrence and/or in predefined time intervals).

For the feature engineering process, this study will draw on previous research to derive a comprehensive overview on smartphone sensed indicators of affect in daily life which can be captured using data collected from smartphone sensors and logs. Where possible, it is also planned to enrich the sensed data with data from other sources (e.g., audio features using the Spotify Web API [2] or Weather data using the OpenWeatherMap API [3]) to ensure that the features reflect the participant's daily behavior and context as comprehensive as possible.

[2] https://developer.spotify.com/documentation/web-api/
[3] https://openweathermap.org/

Analysis Plan

Preprocessing

Inclusion criteria (e.g., criteria for including (1) participants (e.g., Do you only use a subsample?, (2) study days (e.g., only weekdays, certain number of study days), (3) any other criteria concerning data quality (e.g., only days with at least x% of logging data) etc. If you cannot specify these aspects now, please state why.
The following data exclusion criteria will be applied to increase the quality of the analyzed data: - First, participants who did not complete at least five experience sampling (ES) reports of the respective criterion variable (i.e. Valence and Arousal) are excluded. This corresponds to one ES per day on average for the whole two-week sampling phase. Moreover, on the ES level, I will exclude ES reports with an answering time above 900s, which is the maximum time that participants were instructed to spent for filling out the questionnaire. - Second, as some participants might have not used their phone at all for some parts of the ES phase, days of no or nearly no phone usage will be excluded from the study's analyses. Concretely, I will exclude days at which participants had less than ten unlocks of their phone screen or a total usage time of less than 15 minutes. - Third, a minimum number of at least two experience sampling reports per study day is required to enable meaningful predictions for the respective day.

Definition of variables based on smartphone sensing. Please specify your degrees of freedom in variable extraction procedures, e.g., ● *time information (e.g., what does night, daily, weekend exactly mean?)* ● *Aggregation measures (e.g., measures of central tendency/dispersion).* *If you cannot specify these aspects now, please state why.*

Reflected by different sensing modalities, the features included in the predictive modeling will comprise specific behavioral and situational patterns in the domains of (1) communication, (2) mobility, (3) music consumption, (4) smartphone usage, as well as situational characteristics like (5) time, and (6) weather.

To extract interpretable prediction variables, the logged raw phone events are aggregated into sensing features derived from previous research. The prediction variables can be divided into two categories: (1) hourly features and (2) daily features. Hourly features reflect variables which quantify events within one hour before the first answer to the experience sampling questionnaire. Based on the expected prevalence of behavioral manifestations observed in previous large-scale smartphone sensing studies (e.g., Andone et al., 2016; Stachl et al., 2020; Wilcockson et al., 2018), and the SSPS logging frequencies (Schödel & Oldemeier, 2020), a timeframe of 60 minutes before the first ES record will be chosen for all timeframe features. This time window was also identified as suitable in a previous study analyzing the first experience sampling wave of the SSPS data set (currently in preparation for publication by the author of this preregistration together with colleagues).

Concretely, four quantification measures are calculated to aggregate the raw data logs: (1) the minimum, (2) the maximum, (3) the median, and (4) the mean absolute deviation (MAD).

To provide a better understanding of the study's pre-specified features, *table 2* in the *Appendix* displays key terms and framework conditions which were determined for the study's feature engineering process. The exhaustive list of all smartphone sensing features can be found in the additionally uploaded document "*MOOD_preregistration_features.pdf*".

Further preprocessing steps (e.g., transformation of data, handling of missing data/outliers etc.)
Due to technical logging errors, single observations might reach extreme values that do not reflect situation or behavior of the participants. However, the large amount of logging data makes it infeasible to check for outliers manually. Therefore, robust estimators (e.g. median, mean absolute deviation) will be used for feature engineering and a dedicated data

preprocessing procedure will be applied as described in the following. To avoid overoptimistic performance evaluation of the predictive models, pre-processing will be performed within the resampling scheme whenever possible.

- **Data transformation**: Categorical variables (factors) will be re-coded into dummy variables. Moreover, since the Elastic Net models require standardized predictors for regularization, numeric variables will be centered and scaled.
- **Outlier identification**: Extreme outliers defined as values exceeding four standard deviations from the mean will be replaced as missing values. Additionally, features with more than 90% missing values, zero or near-zero variance (10% cut-off), and/or strong correlations with other features ($r > .90$), will be removed following recommendations by Kuhn and Johnson (2013).
- **Missing value imputation**: A median-imputation algorithm will be used. To counteract overfitting, this will be incorporated into the resampling process of the prediction modeling.

Data Analysis

Statistical models
Please specify the statistical model (e.g. t-test, ANOVA, LMM) or algorithms that will be used to test each of your hypotheses. Give all necessary information about model specification (e.g., variables, interactions, planned contrasts) and follow-up analyses. Include model selection criteria (e.g., fit indices), corrections for multiple testing, and tests for statistical violations, if applicable. Please also indicate Inference Criteria (e.g., p-values, effect sizes, performance measures etc.).

Two machine learning regression models are trained on the extracted features to predict self-reported Valence and Arousal scores separately. For both dimensions, separate regression tasks are trained for the momentary mood experience (e.g. including only hourly features) and the daily mood experience (including only daily features).

This study additionally plans to explore the relevance of person aspects in the prediction of individual's mood in daily life. For this purpose, participants' Big Five personality scores (on a facet level) will be additionally added to the four separately estimated prediction models. In summary, eight prediction models are planned to be calculated as shown in *figure 1*.

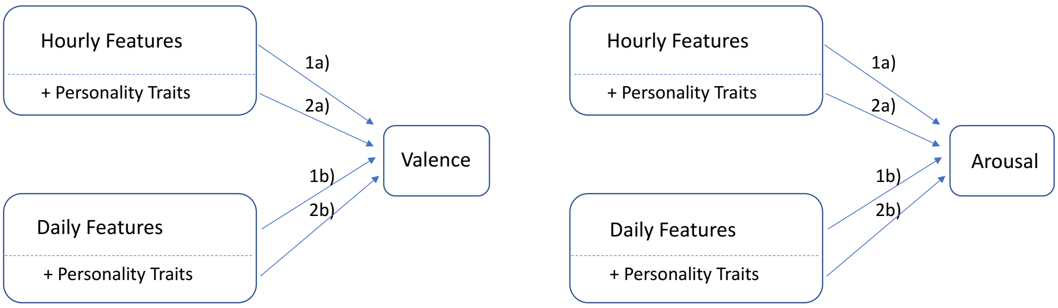

Figure 1. Overview of prediction models

Conducting a statistical benchmarking experiment using the mlr3-package in R (Lang et al., 2019) will enable a systematic comparison of the predictive performance of the Elastic Net regularized regression model (logistic lasso regression; Zou & Hastie, 2005), with a non-linear random forest model (Breimann, 2001) and a featureless baseline model. The baseline model predicts the mean value from the training set for all cases in the test set. Hyperparameters will be tuned in a nested five times repeated ten folds- cross validated resampling scheme. Since the final data incorporates two different data levels (experience sampling (ES) and participant), resampling processes include stratified sampling to avoid oversampling of participants with a larger number of ES reports.

For each model, the predictive performance will be evaluated based on how accurate new (unseen) samples can be predicted. Concretely, cross-validated model fit will be evaluated based on different statistical performance measures, e.g. the root mean squared error (RMSE), the mean absolute error (MAE) and the coefficient of determination (R^2).

This study also aims to apply interpretable machine learning methods by computing feature importance measures for single features as well as the feature groups (i.e. feature categories). Herewith, the predictive power of different smartphone-sensed features for self-reported Valence and Arousal is explored. Moreover, accumulated local effects (ALE plots) and/ or partial dependence plots (PDP) are computed to gain deeper insights in into the direction of specific feature effects.

Planned exploratory analysis (Optional)
To investigate the validation of the predicted Valence and Arousal scores, participant's predicted Valence and Arousal scores will be correlated with their self-reported PANAS scores. Further, descriptive analyses are conducted to investigate if the participants report their mood differently. In other words, some users might use most of the values on the six-point Likert scale, while others might report their affect using only a small portion of the scale (Servia-Rodríguez et al., 2017). Accordingly, it might be considered that the participant is in a positive/ negative (active/ sleepy) mood if he reports values higher/ lower than the median reported by him. In addition to the absolute Valence and Arousal values, I therefore plan to exploratory include the participant's differences to his personal median values as target variables in the prediction models described above.

References

Aichele, S., Rabbitt, P., & Ghisletta, P. (2016). Think fast, feel fine, live long: A 29-year study of cognition, health, and survival in middle-aged and older adults. *Psychological science, 27*(4), 518–529. https://doi.org/10.1177/0956797615626906

Andone, I., Błaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., & Markowetz, A. (2016, September). How age and gender affect smartphone usage. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: adjunct* (pp. 9-12). https://doi.org/10.1145/2968219.2971451

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal, 38*(3), 218. https://doi.org/10.1037/prj0000130

Bergmann, M., Kunz, F., & Schoedel, R. (2021). Sensing psychological situations: Applying machine learning techniques on smartphone-sensed data to predict perceived characteristics of situations in daily life. PsychArchives. https://doi.org/10.23668/PSYCHARCHIVES.4928

Boukhechba, M., Huang, Y., Chow, P., Fua, K., Teachman, B. A., & Barnes, L. E. (2017, September). Monitoring social anxiety from mobility and communication patterns. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 749-753).

Breimann, L. (2001). Random forests. *Machine learning, 45*(1), 5-23.

Cai, L., Boukhechba, M., Wu, C., Chow, P. I., Teachman, B. A., Barnes, L. E., & Gerber, M. S. (2018, September). State affect recognition using smartphone sensing data. In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 120-125).

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., ... & Leow, A. D. (2017, August). Deepmood: modeling mobile phone typing dynamics for mood detection. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 747-755). https://doi.org/10.1145/3097983.3098086

Cheng, H., & Furnham, A. (2003). Personality, self-esteem, and demographic predictions of happiness and depression. *Personality and individual differences, 34*(6), 921-942. https://doi.org/10.1016/S0191-8869(02)00078-8

Ching, C. M., Church, A. T., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., ... & Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: A fiveculture study. *Journal of Research in Personality, 48*, 1-16. https://doi.org/10.1016/j.jrp.2013.10.002

Chow, P. I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L. E., & Teachman, B. A. (2017). Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of medical Internet research, 19*(3), e6820. https://doi: 10.2196/jmir.6820

DeMasi, O., Feygin, S., Dembo, A., Aguilera, A., & Recht, B. (2017). Well-being tracking via smartphone-measured activity and sleep: cohort study. *JMIR mHealth and uHealth, 5*(10), e7820. https://doi.org/10.2196/mhealth.7820

Denissen, J. J., Butalid, L., Penke, L., & van Aken, M. A. (2008). The Effects of Weather on Daily Mood: A Multilevel Approach. *Emotion, 8*(5), 662-667. https://doi.org/10.1037/a0013497

Gable, P. A., Paul, K., Pourtois, G., & Burgdorf, J. (2021). Utilizing electroencephalography (EEG) to investigate positive affect. *Current Opinion in Behavioral Sciences, 39*, 190-195. https://doi.org/10.1016/j.cobeha.2021.03.018

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within-and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality, 69*, 124-138. https://doi.org/10.1016/j.jrp.2016.06.003

Howell, R. T., Kern, M. L., & Lyubomirsky, S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes. *Health Psychology Review, 1*(1), 83-136. https://doi.org/10.1080/17437190701492486

Kashdan, T. B., & Steger, M. F. (2006). Expanding the topography of social anxiety: An experience-sampling assessment of positive emotions, positive events, and emotion suppression. *Psychological Science, 17*(2), 120-128. https://doi.org/10.1111/j.1467-9280.2006.01674.x

Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., ... & Wager, T. (2005). A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological science, 16*(9), 724-731. https://doi.org/10.1111/j.1467-9280.2005.01602.x

Kööts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience. *Journal of individual differences, 32*(2), 74-84. https://doi.org/10.1027/1614-0001/a000037

Kuhn, M. & Johnson, K. (2013). Applied Predictive Modeling. Springer Publishing Company, Inc. https://doi.org/10.1007/978-1-4614-6849-3

Kushlev, K., Dwyer, R., & Dunn, E. W. (2019). The social price of constant connectivity: Smartphones impose subtle costs on well-being. *Current Directions in Psychological Science, 28*(4), 347-352. https://doi.org/10.1177/0963721419847200

Lane, N. D., Mohammod, M., Lin, M., Yang, X., Lu, H., Ali, S., ... & Campbell, A. (2011, May). Bewell: A smartphone application to monitor, model and promote wellbeing. In *5th international ICST conference on pervasive computing technologies for healthcare* (Vol. 10).

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., & Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software, 4*(44), 1903. https://doi.org/10.21105/joss.01903

Lee, H., Cho, A., Jo, Y., & Whang, M. (2017). The Relationships Between Behavioral Patterns and Emotions in Daily Life. In *Advances in Computer Science and Ubiquitous Computing* (pp. 1332-1339). Springer, Singapore.

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013, June). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (pp. 389-402). https://doi.org/10.1145/2462456.2464449

Lu, H., Pan, W., Lane, N. D., Choudhury, T., & Campbell, A. T. (2009). *SoundSense: scalable sound sensing for people-centric applications on mobile phones.* In Proceedings of the 7th international conference on Mobile systems, applications, and services, Kraków, Poland. https://doi.org/10.1145/1555816.1555834

Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012, May). Daily mood assessment based on mobile phone sensing. In *2012 ninth international conference on wearable and implantable body sensor networks* (pp. 142-147). IEEE.

MacLeod, L., Suruliraj, B., Gall, D., Bessenyei, K., Hamm, S., Romkey, I., ... & Meier, S. (2021). A mobile sensing app to monitor youth mental health: observational pilot study. *JMIR mHealth and uHealth, 9(*10), e20638.

Messner, E. M., Sariyska, R., Mayer, B., Montag, C., Kannen, C., Schwerdtfeger, A., & Baumeister, H. (2019). Insights–future implications of passive smartphone sensing in the therapeutic context. *Verhaltenstherapie,* 1-10. https://doi.org/10.1159/000501951

Miranda, D., & Claes, M. (2009). Music listening, coping, peer affiliation and depression in adolescence. *Psychology of music, 37*(2), 215-233. https://doi.org/10.1177/0305735608097245

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the Relationships between Mobility Behaviours and Indicators of Subjective Well–Being Using

Smartphone–Based Experience Sampling and GPS Tracking. *European Journal of Personality, 34*(5), 714-732. https://doi.org/10.1002/per.2262

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). Affect analysis model: Novel rule-based approach to affect sensing from text. *Natural Language Engineering, 17*(1), 95-135. doi:10.1017/S1351324910000239

Nguyen, T., Phung, D., Adams, B, & Venkatesh, S. (2014). Mood sensing from social media texts and its applications. *Knowledge and Information System, 39,* 667–702. https://doi.org/10.1007/s10115-013-0628-8

Ong, A. D., Bergeman, C. S., Bisconti, T. L., & Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *Journal of personality and social psychology, 91*(4), 730. https://doi.org/10.1037/0022-3514.91.4.730

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing, 1*(2), 81-97. https://doi.org/10.1109/T-AFFC.2010.7

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment, 36*(1), 149-161. https://doi.org/10.1027/1015-5759/a000481

Randall, W. M., & Rickard, N. S. (2017). Personal Music Listening. *Music Perception: An Interdisciplinary Journal, 34*(5). https://doi.org/10.1525/mp.2017.34.5.501

Ren, B., Balkind, E. G., Pastro, B., Israel, E., Pizzagalli, D. A., Rahimi-Eichi, H., ... & Webb, C. A. (2022). *Predicting states of elevated negative affect in adolescents from smartphone sensors: A novel personalized machine learning approach.* PsyArXiv. https:://doi.org/ 10.31234/osf.io/zjm4d

Russell, James (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39* (6). 1161–1178.

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality, 69*, 96-101. https://doi.org/10.1016/j.jrp.2016.06.004

Sano, A., Phillips, A. J., Amy, Z. Y., McHill, A. W., Taylor, S., Jaques, N., Czeisler, C. A., Klerman, E. B., & Picard, R. W. (2015). Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In

2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 1–6. doi: 10.1109/BSN.2015.7299420

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., & Barger, L. K. (2018). Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. *J Med Internet Res, 20*(6), e210. doi: 10.2196/jmir.9410

Schoedel, R., Oldemeier, M., Bonauer, L., & Sust, L. (2022). Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset[Data set]. PsychArchives. https://doi.org/10.23668/psycharchives.5362

Servia-Rodríguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017, April). Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 103-112).

Spathis, D., Servia-Rodriguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019, May). Passive mobile sensing and psychological traits for large scale mood prediction. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 272-281).

Steptoe, A., Wardle, J., & Marmot, M. (2005). Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. Proceedings of the National Academy of Sciences, 102(18), 6508-6512. https://doi.org/10.1073/pnas.0409174102

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive affect and psychobiological processes relevant to health. *Journal of personality, 77*(6), 1747-1776. https://doi.org/10.1111/j.1467-6494.2009.00599.x

Stikic, M., Johnson, R. R., Tan, V., & Berka, C. (2014). EEG-based classification of positive and negative affective states. *Brain-Computer Interfaces, 1*(2), 99-112. https://doi.org/10.1080/2326263X.2014.912883

Till, B., Tran, U. S., Voracek, M., & Niederkrotenthaler, T. (2016). Music and Suicidality. *Omega: Journal of Death & Dying, 72*(4). https://doi.org/10.1177/0030222815575284

Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of happiness studies, 9*(3), 449-469. https://doi.org/10.1007/s10902-006-9042-1

Wampfler, R., Klingler, S., Solenthaler, B., Schinazi, V. R., & Gross, M. (2020, April). Affective state prediction based on semi-supervised learning from smartphone touch data. In

Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3313831.3376504

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., ... & Campbell, A. T. (2014, September). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp. 3-14).

Wang, R., Aung, M. S., Abdullah, S., Brian, R., Campbell, A. T., Choudhury, T., ... & Ben-Zeev, D. (2016, September). CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 886-897).

Wang, Z., Ho, S. B., & Cambria, E. (2020). A review of emotion sensing: categorization models and algorithms. *Multimedia Tools and Applications, 79*(47), 35553-35582. https://doi.org/10.1007/s11042-019-08328-z

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining Typical Smartphone Usage: What Data Do We Need? *Cyberpsychology, Behavior, and Social Networking, 21*(6), 395-398. https://doi.org/10.1089/cyber.2017.065

Zhang, X., Li, W., Chen, X., & Lu, S. (2018). Moodexplorer: Towards compound emotion detection via smartphone sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(4), 1-30. https://doi.org/10.1145/3161414

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology), 67*(2), 301-320.

Wichers, M., Smit, A. C., & Snippe, E. (2020). Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: a confirmatory single-subject time-series study. *Journal for Person-Oriented Research, 6*(1), 1. https://doi.org/10.17505/jpor.2020.22042

Appendix

Table 1

Overview of variables investigated in related work on situational and behavioral correlates of mood, emotions, mental health, and wellbeing

Category		Examples	References
Communi-cation	Calls/ text messages	e.g. number of outgoing/ incoming calls, number of outgoing/ incoming text messages, …	Cai et al., 2018; LiKamWa et al., 2013; Lane et al., 2011; Ma et al., 2012; Messner et al., 2019; MacLeod et al., 2021; Sano et al., 2018; Servia-Rodríguez et al., 2019; Wang et al., 2016
	Keyobard logs	e.g. semantic text characteristics, keyboard typing dynamics,…	Cao et al., 2017; Neviarouskaya, et al., 2011; Nguyen et al., 2015; Wang et al., 2020
Mobility	Places	e.g. places visited, home stay, location changes,…	Cai et al., 2018; Ren et al., 2022; Servia-Rodríguez et al., 2019; Sandstrom et al., 2017; Wang et al., 2016
	Movement	e.g. transition time, speed/ acceleration, mobility activity, location entropy/ varaince…	Ben-Zeev et al., 2015; Cai et al., 2018; DeMasi et al., 2017; Lane et al., 2011; Lee et al., 2017; LiKamWa et al., 2013; Ma et al., 2012; MacLeod et al., 2021; Ren et al., 2022; Sano et al., 2018; Servia-Rodríguez et al., 2019; Spathis et al., 2019; Wang et al., 2014; 2016
Music consumption	Listening behavior	e.g. duration of music listening, acoustic characteristics,…	Miranda et al., 2009; Randall & Rickard, 2017; Till et al., 2016; Zhang et al., 2018
Smartphone usage	Apps	e.g. app categories used, duration of app usage,…	LiKamWa et al., 2013; Messner et al., 2019; Wang et al., 2016
	Screen	e.g. screen usage, screen checks,…	Ben-Zeev et al., 2015; DeMasi et al., 2017; Kushlev et al., 2019; Lane et al., 2011; Messner et al., 2019; MacLeod et al., 2021; Ren et al., 2022; Sano et al., 2018;

		Wampfler et al., 2020; Wang et al., 2014; 2016
Time	e.g. morning, evening, night, weekend vs. weekday,…	Cai et al., 2018
Weather	e.g. temperature, wind power, sunlight, humidity, barometric pressure,…	Denissen et al., 2008; Keller et al., 2005; Kööts et al., 2011

Table 2

Name and specification of key terms used in feature description

Key term	Specification
Feature types	
Hourly feature (HF)	Prediction variable based on logging events within a one-hour-timeframe before the experience sampling
Daily feature (DF)	Prediction variable based on logging events within one study day during the experience sampling wave
Features	
Session	Sequence of logging events between a screen unlock and lock event
Check	Sessions with durations of less than 15 seconds (Wilcockson et al., 2018)
App	Mobile applications that are actively used by the user (e.g., no system applications running in the background)
App usage	Sequence of logging events between opening (i.e., moving to foreground of screen) and closing (i.e. moving to background of screen) of an app. If the same app is reopened within 15 seconds after closing it, resulting logging events are assigned to the same usage. If an app is reopened more than 15 seconds after closing it or a new app is opened, resulting logging events are considered a new usage.
Skip	A new song is chosen within a time window of 15 seconds into the current song
Quantifiers	
Min	Minimum value

Max	Maximum value
Average	Measure of central tendency: Median
Variation	Measure of variation: Median absolute deviation (MAD) around the median
Days	
Weekday	Monday, 07:00 – Friday, 18:14
Weekend	Friday, 18:15 – Monday, 06:59
Time	
Morning	7:00 – 10:44 (on Saturdays and Sundays: 9:00 – 12:29)
Noon	10:45 – 14:29 (on Saturdays and Sundays: 12:30 – 15:59)
Afternoon	14:30 – 18:14 (on Saturdays and Sundays: 16:00 – 19:29)
Evening	18:15 - 22:00 (on Saturdays and Sundays: 19:30 – 23:00)

Note. Table adopted from Bergmann et al. (2021) with adjusted feature type descriptions.