Preregistration Protocol: Smartphone Sensing Panel Study - Sensing Psychological Situations

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
Sensing psychological situations: Applying machine learning techniques on smartphone-sensed data to predict perceived characteristics of situations in daily life

Author(s) of the preregistration protocol
Maximilian Bergmann, Fiona Kunz, Ramona Schödel

Date
16.06.2021

Background

Background Information (Optional; Short description of the theoretical background/introduction to research question)
This study is conducted as part of a master thesis at the Department of Psychology of the Ludwig Maximilian University Munich. It investigates whether behavioral and situational data collected via smartphone sensing in daily life can predict individuals' psychological situation. For this purpose, this study applies a machine learning approach to predict individuals' in situ ratings of perceived situational characteristics (DIAMONDS; Rauthmann et al., 2014) based on smartphone sensing data. All data used in this study is retrieved from the Smartphone Sensing Panel Study (SSPS; Schödel & Oldemeier, 2020).

Research question(s)
Can individuals' psychological situation in daily life be predicted from smartphone sensing data?

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.*
As this study uses an exploratory machine learning approach to investigate the research question, we have not formulated any specific hypotheses.

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.*
Self-report Measures Self-report measures analyzed in this study were collected in the first experience sampling wave (27.07.2020 to 09.08.2020) as well as survey one (18.05.2020 to 25.05.2020) and two (13.07.2020 to 19.07.2020) of the SSPS (for a comprehensive description of all measures as well as respective sampling procedures in the SSPS see Schödel & Oldemeier, 2020). In a two-week experience sampling phase, characteristics of participants' psychological situation were assessed between two to four times per day via the German Version of the S8-I Ultra-Brief Measure for the Situational Eight DIAMONDS (S8-I; Rauthmann & Sherman, 2018). The S8-I consists of eight items (one for each DIAMONDS dimension)

for which participants are asked to indicate whether they apply to their current situation. Due to the limited scope of this study, we only include the items for DONS (i.e., Duty: “Work has to be done”, pOsitivity: “The situation is pleasant”, Negativity: “The situation contains negative feelings (e.g., stress, anxiety, guilt, etc.)”, Sociality: “Social interactions are possible or required”) in our data analyses. Diverging from the S8-I by Rauthmann and Sherman (2018), all items were measured on a binary scale (0 = *does not apply*, 1 = *applies*). The binary scale 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 two, participants’ Big-Five personality (Open-Mindedness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) was assessed via the German version of the Big Five Inventory, extra-short form (BFI_2XS; Rammstedt et al., 2020). The BFI_2XS measures the Big Five on the factor-level and comprises 15 items. All items were measured on a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*.

Additionally, participants’ trait affect (positive affect and negative affect) was measured via the German version of the Positive and Negative Affect Schedule (PANAS; Breyer & Bluemke, 2016) in survey two of the SSPS. It consists of ten positive and negative adjectives, which measure positive affect (PA) and negative affect (NA), respectively. Participants rated the degree to which they experience each of the adjectives in general on a 5-point Likert scale from 1 = *not at all to* 5 = *extremely*.

Situational and Behavioral Measures

In the SSPS, a broad range of smartphone sensed situational (e.g., connected Bluetooth devices, location, etc.) and behavioral variables (e.g., calls, texts, app usage, etc.) were assessed via an Android logging app (Android OS version 5 or higher). All collected data are timestamped as they were logged event-based (by occurrence and/or in predefined time-intervals). Depending on the logged event (e.g., calls), it was further specified by additional information (e.g., outgoing). This study will analyze a variety of different event categories logged during the first experience sampling wave of the SSPS (27.07.2020 to 09.08.2020; for a comprehensive list of all event categories and related specifications see Schödel & Oldemeier, 2020).

The situational and behavioral measures that will be utilized in this study are based on events of (1) connectivity (e.g., power cable, headphone status, WiFi, Bluetooth), (2) phone usage (e.g., screen time, calls, texts, apps), (4) mobility/activity (e.g., location, trip, speed, activity state), and (5) timestamp (e.g., day, time).

The variables were selected based on the following process: After identifying situational and behavioral correlates of individuals' psychological situation (i.e., DIAMONDS), manifestations of these correlates in smartphone-sensed data were derived based on previous smartphone sensing literature and theoretical reasoning. For instance, Rauthmann and colleagues (2014) reported that perceived Duty, pOsitivity, Negativity, and Sociality of a situation is associated with the activity of studying or working. This again has been successfully predicted based on users' Bluetooth connectivity data (Chen et al., 2013). Additionally, one could infer such activities based on the users' smartphone usage (e.g., below-average screen time). Consequently, connected Bluetooth devices and screen time were derived as predictor variables. For an exhaustive overview of variables for the prediction modeling see table 2 in the appendix.

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.

To increase the quality of the analyzed data, participants, experience sampling days, single experience samplings and single observations of features will be excluded based on the following criteria.

Participants who reported less than 14 experience samplings of the respective criterion variable (i.e., Duty, pOsitivity, Negativity, Sociality) are removed from the analysis. This represents an average of one experience sampling per day for the complete sampling phase and is equivalent to the minimum requirement for participants to receive compensation.

As some participants might have not used their phone at all for some parts of the experience sampling phase, days of no or nearly no phone usage will be excluded from the study's analyses. This is defined as days at which participants had less than ten unlocks of their phone screen or a total usage time of less than 15 minutes. Note that both measures (i.e., number of screen-unlocks and total usage time) of one day also include the logging events assigned to completion of the ES during that day.

On experience sampling level, all data are excluded in which participants took longer than 900s to complete the experience sampling. This represents the maximum time that participants were instructed to spend filling out the questionnaire.

Due to technical logging-errors, single observations can reach extreme values unrelated to participants' situation or behavior. Thus, distributions of extracted features will be inspected prior to predictive modeling. In case of extremely imbalanced data, outliers will be replaced as missing values. As extreme behavior or situations shall not be excluded from this study, outliers are defined as values exceeding four standard deviations from the mean.

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.

As the collected sensing data consists of timestamped event-data, it will be pre-processed to extract interpretable variables for the applied predictive modeling approach. The prediction variables (or features) can be divided into two categories: (1) status features and (2) timeframe features. Status features are variables that are extracted at the exact timepoint of the experience sampling record. Timeframe features are variables which quantify events within a certain timeframe around the experience sampling. Based on expected situation lengths (e.g., Rauthmann & Sherman, 2016), prevalence of situational and behavioral manifestations (e.g., Andone et al., 2016; Wilcockson et al., 2018), and logging frequencies (Schödel & Oldemeier, 2020), a timeframe of 30 minutes before and after the experience sampling was chosen for all timeframe features. The quantification measures for timeframe features follow previous literature about the relationship of smartphone sensing data and users' situation and behavior (e.g., Chen et al., 2013, 2014; Lu et al., 2009, Min et al., 2013).

To provide a better understanding of the study's pre-specified features, table 1 in the appendix displays key terms and framework conditions which were determined for the study's features. Accordingly, table 2 in the appendix shows all features incorporated in the study's data analyses.

Note that table 2 also displays features that are not included in this thesis. More specifically, features of the event category "data quality" as well as of the subcategories "notifications", "music", "geohash", "displacement", and "distance covered" are not included due to the limited scope of this thesis. However, they were extracted to be included in a research paper aimed for publication. Concerning features comprising categorizations (i.e., device categories, app categories, POI categories, activity categories, manufacturer categories), some of the related categorizations (e.g., app categories) are still developed during this preregistration. Thus, these features cannot be specified for now but will be later in the manuscript. Lastly, several features are contributed by colleagues of the Chair for Psychological Methods and Diagnostics at the Department of Psychology at Ludwig Maximilian University of Munich.

Further preprocessing steps (e.g., transformation of data, handling of missing data/outliers etc.)
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). Finally, missing values of features will be imputed via median imputation. 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.).*
Framed as a binary classification problem, this study will predict reported presence (or absence) of four of the eight DIAMONDS characteristics via situational and behavioral features extracted from smartphone-logs. More specifically, the scores of Duty, pOsitivity, Negativity and Sociality are predicted separately, each in one prediction task. For each model, cross-validated model fit will be evaluated based on how accurate scores of new (unseen) samples can be predicted. Model fit will be evaluated based on the following five performance measures: (1) Accuracy (ACC), (2) Sensitivity (SENS), (3) Specificity (SPEC), (4) F_1-Score (F_1; Chinchor, 1992), and (5) Area under the receiver operating characteristic curve (AUROC). All measures range between 0 and 1 and higher values indicate better model fit. To counteract overestimation of model fit, five times repeated ten-fold cross-validation will be used for predictive modeling. All models are based on data on experience sampling level. As the final data incorporates two levels (L1: experience sampling, L2: participants), resampling processes will comprise stratified sampling and blocking by participant. In a benchmark experiment, predictive ability of a regularized linear model (logistic lasso regression), a random forest, and a featureless model will be evaluated. For all models, this study will use the standard F_1 as a loss function which equally incorporates SENS and the positive predictive value (PPV) in the training process. In case of prediction success (i.e., performance measures of the lasso regression or random forest are above zero and better than the featureless model), variable importance measures are assessed via permutation method. Prediction modeling will be conducted in R using the mlr3-package (Lang et al., 2019).

Planned exploratory analysis (Optional)
If exploratory analyses are possible within the binding time schedule, this thesis will explore the relevance of certain person characteristics in the prediction of individuals' psychological situations in daily life. For this purpose, the prediction models from the main analyses will be enriched with further features representing individuals' traits. More specifically, participants' Big Five personality scores, trait affect scores, and demographic characteristics (i.e., age, gender) will be added to the prediction models. Achieved prediction performance and variable importance will then be analyzed analogously to the main analysis.

References

Andone, I., Błaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., & Markowetz, A. (2016). *How age and gender affect smartphone usage*. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Heidelberg, Germany. https://doi.org/10.1145/2968219.2971451

Breyer, B., & Bluemke, M. (2016). Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis242.

Chen, Z., Chen, Y., Hu, L., Wang, S., Jiang, X., Ma, X., Lane, N. D., & Campbell, A. T. (2014). *ContextSense: Unobtrusive discovery of incremental social context using dynamic bluetooth data.* In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, Washington. https://doi.org/10.1145/2638728.2638801

Chen, Z., Chen, Y., Wang, S., Liu, J., Gao, X., & Campbell, A. T. (2013). *Inferring social contextual behavior from bluetooth traces.* In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, Zurich, Switzerland. https://doi.org/10.1145/2494091.2494176

Chinchor, N. (1992). *MUC-4 evaluation metrics.* In Proceedings of the 4th conference on Message understanding, McLean, Virginia. https://doi.org/10.3115/1072064.1072067

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

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

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

Min, J.-K., Wiese, J., Hong, J. I., & Zimmerman, J. (2013). *Mining smartphone data to classify life-facets of social relationships* Proceedings of the 2013 conference on Computer supported cooperative work, San Antonio, Texas, USA. https://doi.org/10.1145/2441776.2441810

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

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of personality and social psychology, 107*(4), 677-718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. A. (2016). Situation Change: Stability and Change of Situation Variables between and within Persons. *Frontiers in Psychology, 6*(1938). https://doi.org/10.3389/fpsyg.2015.01938

Rauthmann, J. F., & Sherman, R. A. (2018). S8-I - Situational Eight-I - deutsche Fassung [Fragebogen]. In Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Ed.), *Elektronisches Testarchiv (PSYNDEX Tests-Nr. 9007481)*. ZPID. https://doi.org/10.23668/psycharchives.823

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

Appendix

Table 1

Name and specification of key terms used in feature description

Key term	Specification
Feature types	
Status feature	Prediction variable based on one logging event at the timepoint of the experience sampling
Timeframe feature	Prediction variable based on logging events within a one-hour-timeframe around the experience sampling
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)

Table 2

Overview of status- and timeframe features by related event category and subcategory

Event sub-/category	Status feature	Timeframe feature
Connectivity		
Power cable	power cable status is connected	total number of power cable status changes total duration of power cable status connected
Headphones	headphone status is plugged	total number of headphone status changes total duration of headphone status plugged
Flight mode status	flight mode status is on	total number of flight mode status changes total duration of flight mode status on
Bluetooth	Bluetooth status is on and connecting/-ed Bluetooth status is on and disconnected Bluetooth status is off device is connected with "device category"*	total number of Bluetooth status changes total duration of Bluetooth status on and connecting/-ed total duration of Bluetooth status on and disconnected total duration of Bluetooth status off total duration connected with "device category"* total number of different connected device categories
WiFi	WiFi status is on and connecting/-ed WiFi status is on and disconnected WiFi status is off	total number of WiFi status changes total duration of WiFi status on and connecting/-ed total duration of WiFi status on and disconnected total duration of WiFi status off
Phone usage		
Screen time	-	total number of sessions total duration of sessions average duration of sessions min duration of sessions

	-	max duration of sessions variation of duration of sessions total number of checks total duration of checks min duration of checks max duration of checks average duration of checks variation of duration of checks ratio between number of checks and number of sessions ratio between total duration of checks and duration of sessions
Calls	-	total number of calls total duration of call min duration of calls max duration of calls average duration of calls variation of duration of calls
	-	total number of outgoing calls total duration of outgoing call min duration of outgoing call max duration of outgoing calls average duration of outgoing calls variation of duration of outgoing calls
	-	total number of incoming calls total duration of incoming calls min duration of ringing of incoming calls max duration of ringing of incoming calls average duration of ringing of incoming calls variation of ringing of incoming calls min duration of incoming calls max duration of incoming calls average duration of incoming calls variation of duration of incoming calls

	-	total number of missed calls total duration of ringing of missed calls min duration of ringing of missed calls max duration of ringing of missed calls average duration of ringing of missed calls variation of duration of ringing of missed calls
	-	total number of rejected calls total duration of ringing of missed calls min duration of ringing of rejected calls max duration of ringing of rejected calls average duration of ringing of rejected calls variation of duration of ringing of rejected calls
Texts	-	total number of texts min length of texts max length of texts average length of texts variation of length of texts
	-	total number of outgoing texts min length of outgoing texts max length of outgoing texts average length of outgoing texts variation of length of outgoing texts
	-	total number of incoming texts min length of incoming texts max length of incoming texts average length of incoming texts variation of length of incoming texts
Apps	-	total number of app usages total duration of app usages total number of unique apps used

		min total number of usages per app max total number of usages per app average total number of usages per app variation of total number of usages per app min total usage duration per app max total usage duration per app average total usage duration per app variation of total usage duration per app
	-	total number of different app categories used min number of total usages per app category max number of total usages per app category average number of total usages per app category variation of number of total usages per app category min total usage duration per app category max total usage duration per app category average total usage duration per app category variation of total usage duration per app category total number of usages of "app category" apps* total usage duration of "app category" apps*
Music	music is currently listened to Level of "Spotify audio feature category" of currently listened song[+]	total number of songs listened to total duration of songs listened to total number of skips min level of "Spotify audio feature category" of listened songs[+] max level of "Spotify audio feature category" of listened songs[+] average level of "Spotify audio feature category" of listened songs[+] variation of level of "Spotify audio feature category" of listened songs[+]
Notifications	-	total latency of notification caused app usage min latency of notification caused app usage max latency of notification caused app usage average latency of notification caused app usage variation of latency of notification caused app usage

Mobility/activity			
	Location	current location is at home current distance from home	total time spent at home min distance from home max distance from home average distance from home variation of distance from home
		current location is at work current distance from work	total time spent at work min distance from work max distance from work average distance from work variation of distance from work
		current location is at "POI category"* density (per km^2) of inhabitants of POI City number of inhabitants of POI City density (per km^2) of inhabitants of POI District number of inhabitants of POI District	-
	Altitude	current altitude (m above sea level)	min altitude max altitude average of altitude variation of altitude total altitude change altitude positive change altitude negative change
	GeoHash	total number of visits of current GeoHash (within complete experience sampling phase) total time spent at current GeoHash (within complete experience sampling phase)	total number of different visited GeoHashs min time spent per visited GeoHash max time spent per visited GeoHash average time spent per visited GeoHash variation of time spent per visited GeoHash

Displacement	-	standard deviation of displacements radius of gyration location variance
Distance covered	-	total distance covered spatial coverage by convex hull maximum distance between two locations
Trip	participant is on an identified trip (vs. stay) total time spent at current trip (within complete experience sampling phase)	total number of different trips identified min time spent per identified trip max time spent per labeled trip average time spent per labeled trip variation time spent per labeled trip
Speed	current speed (m/s)	min speed max speed average of speed variation of speed total speed change total time spent in transit
Activity state	Probability of the current activity being “activity category”*	min probability of “activity category”* max probability of “activity category”* average probability of “activity category”* variation of probability of “activity category”*
Timestamp		
Time	timepoint of experience sampling is at morning timepoint of experience sampling is at noon timepoint of experience sampling is at afternoon timepoint of experience sampling is at evening	-
Day	current timestamp is at the weekend (vs. weekday)	-

Data quality		
Manufacturer	manufacturer of smartphone is "manufacturer category"*	-
Battery mode	battery saving status is on	total duration of battery saving status on total number of battery saving status changes
Location accuracy	current location accuracy	min location accuracy max location accuracy average of location accuracy variation of location accuracy
Altitude accuracy	current altitude accuracy	min altitude accuracy max altitude accuracy average of altitude accuracy variation of altitude accuracy

Note.

POI = Point of interest;

All categorial status features are dummy coded with 1 = *yes*, 0 = *no,* if not specified differently (e.g., weekend vs. weekday);

Timeframe features are extracted within a timeframe of 30 minutes before and after the experience sampling was started by the participant. Consequently, quantifiers (e.g., total, ratio, min, max, mean, variation) of timeframe features are based on logging data within this one-hour-timeframe;

* indicates that this feature will be extracted for each category of the related categorization (i.e., device categories, app categories, POI categories, activity categories, manufacturer categories). As some categorizations are still under development, they will be specified later in the manuscript;

+ indicates that this feature will be extracted for each category of available Spotify audio features (i.e., danceability, acousticness, energy, instrumentalism, liveness, loudness, speechiness, tempo, valence);

Features of the event category “data quality” as well as of the subcategories “notifications”, “music”, “geohash”, “container”, “displacement”, and “distance covered” will not be included in the study’s data analyses. However, they are extracted to be included in a research paper aimed for publication;

Features of the subcategories “calls”, “texts”, and “apps” are extracted by the thesis’ author. All further features have been extracted in previous work or are still under development by colleagues of the Chair for Psychological Methods and Diagnostics at the Department of Psychology at Ludwig Maximilian University of Munich.