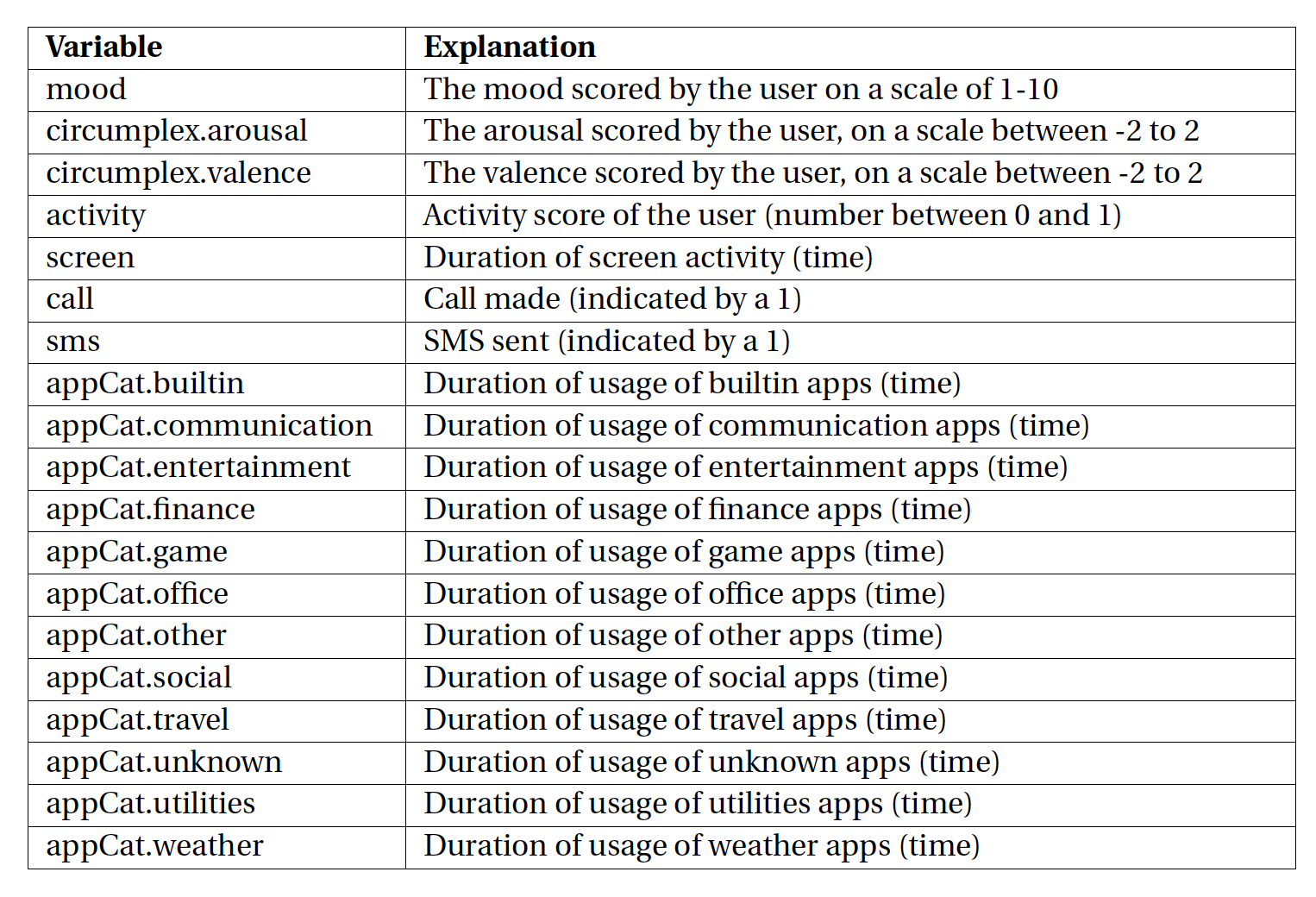
**Minutes 2023-04-11**



**To discuss**

* “call” values are all 1, should we assume 0 as not responded or no call that day? Currently treating them as no call
* A patient might input multiple mood on a day, how should we treat it? currently we take average
* Remove all non-appCat features (including original and derived) with more than X% missing values?
  + X = 80 🡪 remove activity, screen
  + X = 75 🡪 remove call, arousal, activity, screen
  + X = 70 🡪 remove call, arousal, valence, mood, activity, screen (basically everything ☹ )
* Three datasets:
  + Base dataset with original features and missing values handled
  + Normalized datasets with original features and missing values handled
  + Derived, normalized datasets with averaged numerical features and transformed features (see feature engineering for feature)
    - 2023-04-11 dataset: normalized, taking average metric over 5 days of mood, circumplex.arousal, circumplex.valence, activity, screen, call, sms, and appCat aggregated attributes. For appCat attribute, we summarised by category: work (finance, office), recreation (communication – social – entertainment – game), utils (builtin – utilities – weather – travel), others (other – unknown)

**Idea for new features/feature engineering**

* Difference between mood and average mood of last 5 days
* New time features on weekday – weekend, or on time of the day (morning, afternoon, evening)
* Scaling: linear or exponential

**Todos for feature engineering**

* Correlation test between mood at time t + 1 and attribute at time t: , with hyperparameter threshold
* Probably some t-testing