IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 8, 2019, accepted August 23, 2019, date of publication September 6, 2019, date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939960

Sequential Pattern Mining Suggests Wellbeing Supportive Behaviors

MUHAMMAD J. ALIBASA[1], **RAFAEL A. CALVO**[2,3], **(Senior Member, IEEE), AND KALINA YACEF**[1]

[1]School of Computer Science, The University of Sydney 2006, Australia
[2]Dyson School of Design Engineering, Imperial College London SW7 2AZ, U.K.
[3]School of Electrical and Information Engineering, The University of Sydney 2006, Australia

Corresponding author: Muhammad J. Alibasa (johan.alibasa@sydney.edu.au)

The work of R. A. Calvo was supported in part by the Australian Research Council Future Fellowship under Grant FT140100824. The work of M. J. Alibasa was supported by the Indonesian Endowment Fund for Education (LPDP).

ABSTRACT Amidst the headlines about the attention economy and the possible impacts of screen time, research investigating the complex relationship between digital technology usage and wellbeing has gained urgency. Researchers generally use a combination of surveys and automatic tracking tools to gather time and frequency of technology use. However, the focus of data analysis has been on measuring duration and frequency of usage rather than exploring behavioral patterns, possibly better indicators of mood states or stress levels. We propose a methodology for detecting behavioural patterns from digital footprints using a sequence pattern mining algorithm, and using these as features for predicting mood. Results show that our method can be used to analyze the relationship between digital usage and mood, and predict the latter with an accuracy of 80%, significantly above the baseline (71.1%). This method provides another angle to investigate digital technology usage in wellbeing-related research.

INDEX TERMS Screen time, app usage, digital footprints, affective computing, mood detection.

I. INTRODUCTION

Computer and smartphones have become part of everyday life and, as we spend more time in front of screens, concerns about their impact on health have reached fever pitch. Our digital experiences impact our moods, and researchers have argued that they can even support or hinder our mental health and wellbeing [1]. For over two decades, research has shown that digital activities can elicit positive emotions and moods, e.g., listening to music while at work [2] or sharing photos [3]. However, their relationships are complex. While in some studies, digital technology usage seemed to boost wellbeing, many other studies show that some of them may impair wellbeing. Email, for example, has shown to induce higher stress when used frequently [4], as have other workplace technologies [5]. On the other hand, other research suggests that technologies, like computer games can reduce stress at work [6]. Studies often seem to contradict each other, for example, time spent on Facebook has shown to worsen mood (compared to neutral browsing) [7], but to be correlated with positive moods at the end of the day [8], or correlated to social anxiety symptoms [9]. In the end, the relationship between digital technology use and wellbeing is still unclear and sometimes contradictory. Today's research may just be the tip of an iceberg filled with psychological and sociological research questions.

Past studies investigated this relationship by combining mood survey data with aggregated usage data of applications and websites. Earlier studies used self-reports to collect data about both behaviour and mental states, but this approach can be quite inaccurate. The data in today's studies usually comprise general digital frequency usage, e.g., rating on the frequency of engagement with Facebook activities [9], and increasingly, objective measurements of digital technology usage. In recent times, researchers have started to use automatic digital tracking tools such as RescueTime [10] and Kidlogger [11]. However, they only utilised standard measurements such as total duration or total frequency of using particular applications [12], [13], and total application-switches [8], whilst potentially missing important information about how the activities occur and whether there are specific patterns of usage that can indicate negative mood or distress. For example, two people may have spent overall the same amount of time on Facebook and MS Excel, but

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng.

with very different usage patterns: one may have used them sequentially and continuously whereas the second may have switched constantly between the two applications. It is possible that these two patterns may be associated with different moods, for example. In our work, we aim to explore whether recurrent and contrasting digital activity patterns can be used in a new type of exploration of how designers can support wellbeing.

Companies already collect digital behaviour data, often to explore the impact on wellbeing, such as Google Digital Wellbeing,[1] Apple Screen Time and Microsoft MyAnalytics. If such data can be used to predict mood, a mood detection model could be built to predict mood unobtrusively. Prolonged negative mood and lack of positive moods are common predictors of mental health issues [14] and the positive balance (more positive and fewer negative) is itself one definition of wellbeing (*i.e., hedonic*) [15]. Unobtrusiveness is also an important aspect in affective computing [16] and less-intrusive sensors have better user acceptance [17]. This finding opens up opportunities to predict mood continuously while users use their computer or smartphone.

Our paper makes two contributions: (1) a methodology for identifying digital technology usage patterns associated with mood and (2) a model for mood prediction, based on these patterns and duration-derived features. This paper also confirms two hypotheses. The first hypothesis is that using sequence patterns will reveal patterns that involve commonly used applications and that are more frequent prior to positive moods than prior to negative moods, and vice-versa. To verify this hypothesis, we designed a preprocessing technique with a parameterised bucket-window model to represent activity vectors. We then used a Generalized Sequence Pattern algorithm (GSP, [18]) to find interesting patterns and contrasted their occurrences between positive and negative moods. We fine-tuned our parameters and preprocessing steps which will be appropriate and useful. The results in Section V-A will answer the first hypothesis. Our second hypothesis is that sequence patterns are good predictors to detect mood. To confirm this, we use the most frequent digital patterns as mood detection features and discuss the performance of our classifiers in Section V-B. We evaluate their performance against several mood detection models from existing literature that use duration-derived features. Lastly, we discuss our findings and demonstrate how our method can be applied to future work investigating the impact of digital technology use on wellbeing.

II. RELATED WORK

There are numerous definitions of wellbeing, including clinical (where it is defined as the absence of illness) or *eduaimonic* where it is defined as in broader terms of a 'life well lived'. In this paper we will take the narrow definition of *hedonic* wellbeing that refers to the balance of positive and negative emotions. Despite its limitations, this is the approach most often used by economists like Kahneman *et al.* [19], and affective computing studies.

Affective computing brings together human computer interaction and psychology and has provided evidence of the relationships between digital experiences and moods/emotions. Some of the connections might seem obvious (e.g., watching a movie) but some less so (doing email). It is important to study these connections in specific contexts given the complexity of how wellbeing relates to technology experience. To investigate these relations an obvious starting point is data on the time spent on an application or website, part of what is known as digital footprints. This data can be collected using commercial tools (e.g., RescueTime[2]). Researchers may develop their own digital tracking app or use available commercial tracking applications. However, by large, the data used for analysis has been high level measures not using the rich and detailed information available in digital traces, their timing and their patterns.

Previous studies used the total duration of particular applications or websites to analyze the relationship between digital technology usage and mood. A study calculated the number of seconds spent on email client app [13] to find the relationship between email usage and mood. The study used the total email duration for each day and ran a statistical analysis to obtain correlation results. In another study, Mark *et al.* [8] also used total duration data for Facebook, email and calendar usage. The study also included the number of application switches and switches between documents in addition to the total duration data. However, they did not analyze the application-switch patterns even though the study discussed the relevance of task switching. In particular they did not explore whether their participants had a tendency to switch between particular applications when they reported negative moods.

Engagement duration data have been shown to be useful as features to predict mood. A recent study [20] used the duration data from several digital categories for detecting two mood labels (positive and negative). Their mood prediction model shows an accuracy between 81-82%. However, these results were from the dataset where the digital activity records were complete or near complete. For instance, they required 30-minute data in 30-minute window dataset and 55-minute data in 60-minute window dataset. Therefore, the mood prediction model would perform worse when there are missing data, i.e., users did not use the computer/smartphone for some period of time. Also, this study did not consider the digital patterns that might help detect mood. Though, this paper indicates that the mood prediction may improve when combined with digital usage patterns.

Beside mood detection, digital footprint data has been used by Ferdous et al. to predict stress levels based on smartphone application usage [12]. These authors extracted features including total duration and total frequency of various application categories, such as social networking service,

[1] https://wellbeing.google/

[2] https://www.rescuetime.com

entertainment, utility, browser, and game. By utilising these two measurements, they were able to develop a stress detection model with an average accuracy of 75% and precision of 85.7% while using user-centric model. However, their model could only yield 54% accuracy (10-fold cross-validation) when using a generic model (combining datasets from all participants).

Digital footprints have also been used to predict Big Five personality types [21]. In this study, Grover et al. compared two sources (computer and smartphone) and proposed digital trace model called temporal patterns. This model has five metrics: dispersion of technology usage, social media ratio, 5-minute window social media ratio, 'rhythm' of usage, and evening usage ratio. However, these metrics only use simple duration measures including device (computer and smartphone) usage duration and social media (e.g., Facebook) usage duration. Therefore, this temporal pattern mainly focuses on the routines on how people with peculiar personalities use the device and social media. A new metric that considers the application switch pattern can also provide more perspective or information, for example extroverted people might have a tendency to switch between instant messaging and social media more often than others.

Brdiczka et al. used a different type of behavioural temporal pattern [22] to investigate stress factors (productivity, autonomy, and workload) on routine tasks [23]. Using this pattern detection algorithm, they analyze the number of repeated occurrences (temporal patterns) of application, document, and email events. From this algorithm, they obtained the number of unique significant temporal patterns ($p<.05$) and the average of minimal duration of the temporal patterns. The study suggested that these measures were able to indicate workplace stress factors. However, they could not show specific application-switch patterns that were often repeated when their participants were under high-stress level.

Most of the previous studies combine the total duration of digital technology use for each day. However, it is possible that mood or stress level can change within a day depending on the contexts and events [24]. For example, a student might be in a negative mood before the deadline while using writing software, but the mood can change to positive after submitting the assignment and watching YouTube videos. Therefore, we believe it is important to focus on shorter time windows when analysing mood or stress patterns.

These gaps led us to explore a different approach for analysing the relationship between mood and digital activity that analyses users' temporal digital footprints and their switching patterns occurring at the time they report their mood.

III. DATA COLLECTION

Our study collected data from participants through two applications: RescueTime[2] and MindGauge.[3] RescueTime is a commercial time management application that tracks the total time spent on applications and websites (from both computer and/or smartphone). Each RescueTime record row includes timestamp, duration (time spent), application/website name, category name and productivity value. In our analysis, we used timestamp, duration and category name. We selected category name instead of application/website name as we want to analyze less granular data, e.g., Facebook and Twitter are treated as "Social Networking", Gmail and Outlook as "Email", digital libraries websites and any learning management systems (LMS) as "Reference and Learning".

The second application used for our study, MindGauge, is a mobile application designed to help users understand their digital technology usage in relation to wellbeing and work. This mobile application provides a simple mood widget so users can log mood self-reports. The mood widget contains a question "How are you feeling now?" with five answer options (great, good, neutral, bad, and awful). These five labels were then mapped into two labels (positive and negative mood) and these two labels are treated as our ground truth for the analysis. Beside the mood label, each self-report has other values, such as user ID and timestamp. This timestamp is used for aggregating preceded digital activity data prior to mood self-report event.

Participants were recruited via Google Ads campaign and snowball sampling and consented as specified in a study protocol (2016/855) approved by the Human Research Ethics Committee at the University of Sydney. Between January 2018 until April 2019, 236,008 RescueTime data rows were collected over 1,106 unique days (mean≈15 days per user). The participants also entered 1,244 mood self-reports with 707 reports labelled as positive mood and 537 reports as negative mood (including neutral). We used two classes (positive, negative) to maximise possible accuracy and meaningful labels. Using more labels or classes would lower accuracy and be susceptible to imbalance dataset issues. These data were collected from a total of 72 participants who consented to share their MindGauge and RescueTime data (29 male users, 35 female users, and 8 users who preferred not to provide their gender information). Their age ranges between 18 to 58 with average 30.9 years old (13 users did not provide this information).

MindGauge uses an experience sampling technique that asks users' mood or feeling two times at random times on working days (9AM to 6PM) and once at the end of the day (8PM). Beside this sampling, MindGauge also asks end-of-day mood sampling, as in past studies [11], [13]. We analysed mood fluctuations during the day. From a 5-likert scale, the standard deviation average from experience sampling is 0.24. After comparing the average of experience sampling mood reports and the end-of-day mood reports, we obtained a Mean Square Error value of 0.58. Further, Wilcoxon test was performed to compare the data distribution between the average moods from experience sampling and the end-of-day moods. The test result shows that their data distributions are significantly different ($p=0.0098$). These results indicate that mood can fluctuate during the day depending on events and

[3]https://positivecomputing.bitbucket.io/mindgauge/

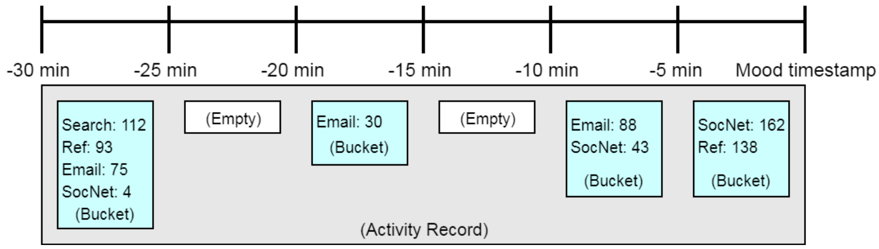

FIGURE 1. Data bucket and activity records example for 30 minutes prior to self-report timestamp.

contexts, as found by Isen *et al.* [24]. Therefore, this study focuses on data over relatively short time windows (30, 60, 90 minutes) occurring just prior mood self-reports to predict mood.

IV. METHOD

We sought to extract the patterns of digital activity records (from RescueTime) in specific windows of time preceding the timestamped self-reported moods (from MindGauge). The key aspect of our method consists in first pre-processing the raw digital activity records into meaningful set of sequences, and then detecting the top sequential patterns preceding positive and negative moods, respectively. We used generalized sequential pattern (GSP) [18] which was successfully used for detecting frequent sequences from digital footprints (e.g., in collaborative learning data [25]), and in particular SPMF, an open-source data mining library [26], to run the GSP algorithm. From the GSP results, we use the most frequent digital patterns found as features in the mood detection model. Next we introduce several terms that help make the preprocessing steps clearer and easier to follow.

A. DATASET PROPERTIES AND DESCRIPTIONS

RescueTime data were grouped into *activity data buckets* that each includes time spent on each digital category over five minute time windows. We chose a five-minute length to be consistent with previous studies [8], [21]. One activity data bucket can contain information, such as the following indicating that the user, over the 5 minute period considered, spent 162 seconds on Social Networking and 138 seconds on Reference and Learning:

```
[
 {'category': 'Social Networking', 'timeSpent': 162},
 {'category': 'Reference & Learning', 'timeSpent': 138}
].
```

The 'timeSpent' values are in second unit and 'category' values contain the category name. Activity data buckets are indicated by the mention (Bucket) in Figure 1.

To generate sequences for GSP, we considered the activity data buckets that occurred within a time window immediately preceding a mood self-report timestamp. As described in section 2, mood can change frequently depending on events and contexts, so we investigated various time-window lengths in our study: 30 minutes, 60 minutes, 90 minutes and 120 minutes. This time-window prior to a mood self-report timestamp is represented by a linked list of activity data buckets and is referred to as an *activity record*, and will be the sequence input for the GSP algorithm.

The number of data buckets per activity records varies for each activity record, depending on the length of the time window considered and possible interruptions in the activity resulting in empty buckets. A 30-minute activity record therefore contains 0 to 6 data buckets (or up to 12, 18 or 24 respectively for 60, 90, 120-minute ones). For instance, if a user makes a self-report after using their computer for 20 minutes, with two 5-minute interruptions where they did not use their computer, then the corresponding activity record would only contain four data buckets (≈20 minutes data) within 30 minutes prior to the self-report timestamp. Figure 1 illustrates this example of such 30-minute activity record.

Some mood self-reports were not preceded by much digital tracking data within the considered time window. This might happen as users can enter mood self-reports anytime even though they did not use their computer or smartphone prior to the self-report. Since we are trying to predict mood based on digital activity, these occurrences are not useful for our purpose. To filter these out, we introduce a new term called *bucket-bound* (N_{bound}) which indicates the minimum number of activity data bucket(s) required for an activity record to be used for analysis. For example, using the previous example (Figure 1), if we have N_{bound} value set to 3 then we can include this activity record as a sequence since the record has 4 data buckets. However, we would not include this activity record as a sequence if the N_{bound} value was set to 5. In our analysis, we experimented with various N_{bound} size and observed the impact on the results.

1) PREPROCESSING

We generated activity records for every mood self-report from each user and classified them into two labels, positive and negative mood. We iterated each user and obtained both positive and negative labelled self-report timestamps. Afterwards, we searched for data buckets from RescueTime data within X minutes prior to timestamp ($X \in \{30,60,90,120\}$). Each data bucket contains digital categories (e.g., 'Email' or 'Social Networking') and the corresponding time spent values. We excluded periods spent in a digital category for less than 10 seconds as it is considered not long enough to be significant. Users are not entirely engaged in this digital activity if they use it less than 10 seconds, so we removed these outliers from our data buckets. Finally, the encoded categories in each data bucket were sorted by activity ID mapping as it is mandatory to sort the content of transactions before serving as input to the GSP algorithm.

2) RUNNING GSP ALGORITHM

After every data bucket has been preprocessed, we generated sequences based on data buckets of each activity record. We then checked whether the data bucket size of each activity record is greater or equal to the bucket-bound size (N_{bound}) If the record met the requirement, then it is included in the positive or negative mood array, depending on the reported mood. Finally, we use these sequences as inputs for GSP algorithm on SPMF.

SMPF only accepts text file inputs with specific formats to run the GSP algorithm. Let us consider the activity record as shown in Figure 1, where we have 4 data buckets. From the first data bucket, we have 'Search', 'Reference & Learning', and 'Email', while 'Social Networking' category was removed because the duration is less than 10 seconds. For example, by using activity ID mapping, 'Search' category is mapped into 11, 'Reference & Learning' into 12, 'Email' into 13, and 'Social Networking' into 14. In this example, we will have the following sequence: *<('Search', 'Reference & Learning', 'Email'), ('Email'), ('Email', 'Social Networking'), ('Social Networking', 'Reference & Learning')>* mapped and sorted into *<(11,12,13),(13),(13,14),(12,14)>*. For better readability, we will present results in short hand (e.g., "Gms" for Games etc.) rather than integers. Each item in data bucket is separated with a space " ". Afterwards, each data bucket in each activity record is separated with a mark "−1". Lastly, every activity record (each line of the text file) is marked by "−1 −2". If we use the previous example *<(11, 12, 13), (13), (13, 14), (12, 14)>*, then we will have "*11 12 13 −1 13 13, 14 −1 12 14 −1 −2*". We generated files with combinations of each mood type (negative or positive), activity record time window size and N_{bound} size.

We set the minimum support threshold to 10% to generate enough patterns for analysis of this study. Our review of the literature suggests that there is no gold standard in terms of minimum support value for digital behaviour data since no study has investigated the sequence using GSP. In this paper, only the top 10 of each mood were presented. Also, there is no sequence-length limit for generating the patterns so patterns with more than two sequences can be obtained.

Finally, we removed outputs that include undefined categories, such as 'Other', 'Utilities', and 'Uncategorized' since they are too general and therefore do not provide useful information.

We ran the GSP algorithm with activity record time window values: 30, 60, 90, 120 and N_{bound} values: 1, 2, 3, 4, 5, 6 to compare the results of each parameters.

TABLE 1. Popular examples of each digital category.

Category	Application/Website Examples
Reference & Learning	Adobe Acrobat, Preview, Wikipedia, Learning management system (LMS)
Social Network	Facebook, Instagram, Twitter
Search	Google, Bing
Email	Gmail, Outlook, Yahoo Mail
Writing	Google Docs, Microsoft Word, Overleaf
Business	MS Excel, Google Drive, Fastway, Tableau
Games	IGN, Steam, Hearthstone, Dota2, Fortnite
News & Opinion	NYTimes, Forbes, Reddit, Medium
Entertainment	Spotify, Audible, Webtoons, Glamour
Browsers	Google Chrome, Firefox, Internet Explorer
Instant Messaging	Whatsapp, Slack, Wechat, Telegram
Video	Youtube, Netflix, Twitch, Bilibili

B. MOOD DETECTION MODELS

We built mood detection models that use the frequent digital activity patterns as features, i.e., the presence of particular frequent digital patterns in the activity records. We will discuss the selected features of these models in section V-B after presenting the GSP results. In addition to using ZeroR classifier as a baseline, we implemented two more sophisticated baseline mood detection models based on existing literature. Both use duration-related features. We then created two models that use the frequent patterns found from methods described in Section IV-A as features for mood prediction.

1) RAW DURATION (RD) MODEL

The first model was built using features from [12] that used five categories of digital footprints to detect stress: social networking service, entertainment, utility, browser, and game applications. Both duration and frequency were extracted from each category as features. We mapped RescueTime categories into these five categories based on the examples provided in their study. The model also used the number of unique application/website names as the feature.

2) DURATION PERCENTAGE (DP) MODEL

The second baseline model also used duration-related features based on the mood prediction study by [20]. Compared to the RD model, this model uses duration percentage of selected RescueTime categories within particular time-windows. We will call this model as Duration Percentage (DP) model. We expect that the DP model performance will be better than the RD model since the duration percentage value is adding some normalisation. A total of twelve popular digital categories were selected as machine learning features (Table 1) for the DP model. While the RD model used the number of unique application/website names, the DP model uses the number of unique categories to make it less granular and consistent with past studies. Lastly, the model also includes a new feature, the RescueTime productivity score [27].

3) DIGITAL SEQUENCE PATTERNS (SP) MODEL

This model utilises solely digital patterns as mood prediction features. As previous studies have not considered them to

TABLE 2. Top 10 categories in 30-minute window data for negative mood.

[N_{bound}=1, N=191] category	sup (%)	[N_{bound}=2, N=158] category	sup (%)	[N_{bound}=3, N=130] category	sup (%)	[N_{bound}=4, N=106] category	sup (%)
Search (Srch)	39	Search (Srch)	42	Search (Srch)	47	Search (Srch)	50
Reference (Ref)	34	Reference (Ref)	39	Reference (Ref)	45	Reference (Ref)	48
Email (Em)	28	Email (Em)	30	Social network (SN)	33	Social network (SN)	37
Social network (SN)	26	Social network (SN)	30	Email (Em)	32	Email (Em)	35
Browsers (Brw)	22	Browsers (Brw)	25	Browsers (Brw)	29	Business (Bsn)	31
Business (Bsn)	19	Business (Bsn)	23	Business (Bsn)	26	Browsers (Brw)	29
Games (Gms)	18	News & Opinion (Nw)	22	News & Opinion (Nw)	24	News & Opinion (Nw)	26
News (Nw)	18	Games (Gms)	21	Entertainment (Ent)	22	Entertainment (Ent)	25
Entertainment (Ent)	16	Entertainment (Ent)	20	Games (Gms)	21	Writing (Wrt)	25
Writing (Wrt)	15	Writing (Wrt)	18	Writing (Wrt)	20	Games (Gms)	19

predict mood, this model can explore the initial feasibility. The digital patterns used for prediction features were selected from the most sequence patterns obtained. These features will be discussed further in the next section.

4) COMBINATION OF DURATION AND SEQUENCE (DP+SP) MODEL

This mood detection model uses both digital sequence patterns (SP) and features from DP model. Duration features provide information about the relative importance of each digital category while digital footprints provide information about switching patterns. Our hypothesis was that these features will complement each other and yield higher mood prediction accuracy.

C. CLASSIFIERS

We tested several classifiers including SVM, Neural Network, Linear Regression, Random Forest and Gradient Boosting. Random Forest and Gradient Boosting classifiers, yielded higher accuracy compared to the others, and this paper will only describe the results from these classifiers. Both RF and GB are ensemble learning (the wisdom of crowds) methods, which are models that combine different individual models and tend to have less bias and less variance.

Five models were tested on the datasets. The datasets were selected from 60-minute activity record windows with $N_{bound} = 6$. We set N_{bound} size to 50% of the maximum bucket size for these activity record time windows. From each user, the last 20% data were used for testing as shown in Fig. 2. The test sets were not included to train the model and only be used for the final evaluation. The first 80% datasets from each user were used for training to build the mood prediction model and to find the best hyperparameter for each model. GridSearchCV with 10-fold cross-validation method was used to obtain the best model parameters that yield the best cross validation score (average accuracy). From the test datasets, following metrics were used to evaluate our models when to predict positive vs. negative mood: accuracy, recall (true positive rate), specificity (true negative rate), precision, and F1 score.

V. RESULTS

A. GENERALIZED SEQUENTIAL PATTERN

The GSP algorithm produced sequences of various lengths with support values. Let us first look at the frequency

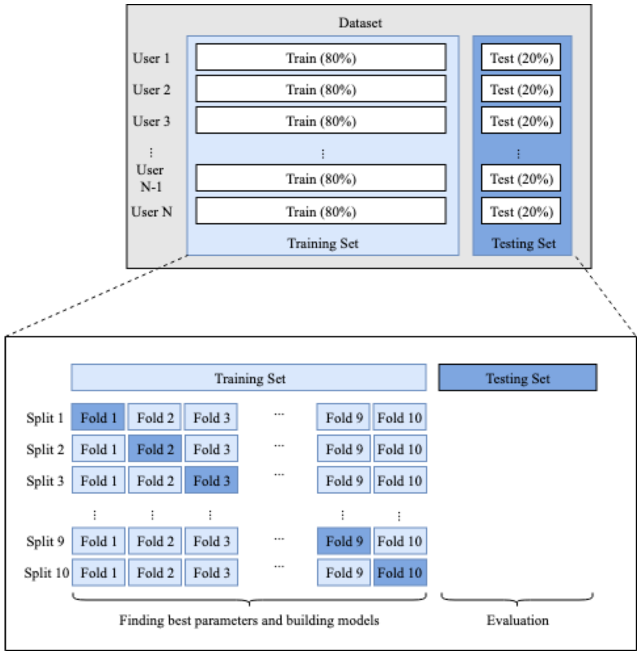

FIGURE 2. Training, model selection, and evaluation.

of activity categories, or unary sequences. These are not sequences per se but we simply use them here to see the frequency distribution. These are shown in Table 2 (for negative moods) and Table 3 (for positive moods). These tables illustrate the top ten categories for each bucket-bound size (1,2,3,4) while using 30-minute time window. They can inform which categories are common for each mood in each parameter set. From the header of these tables, we notice that the number of sequence inputs (N) is decreasing as the bound-bucket size increases. The reason for this trend is that there are activity records containing few data buckets. For instance, in Table 2, there are 191 sequence inputs with $N_{bound} = 1$, but only 158 sequence inputs with $N_{bound} = 2$. It means that 33 (≈17%) sequence inputs have only one transaction (from one activity data bucket). As the bucket-bound size increases, the support values are also rising since sequence inputs with less transactions are removed.

The top ten categories tables show several interesting findings. Negative mood label has ‘Search’ (Srch) category as the most commonly used digital category, while positive has ‘Social Networking’ (SN) category. Both mood labels have

TABLE 3. Top 10 categories in 30-minute window data for positive mood.

[N_{bound}=1, N=278]		[N_{bound}=2, N=237]		[N_{bound}=3, N=197]		[N_{bound}=4, N=167]	
category	sup (%)	category	sup (%)	category	sup (%)	category	sup (%)
Social network (SN)	41	Social network (SN)	44	Social network (SN)	48	Social network (SN)	50
Email (Em)	39	Email (Em)	41	Email (Em)	45	Reference (Ref)	47
Reference (Ref)	36	Reference (Ref)	41	Reference (Ref)	44	Email (Em)	46
Writing (Wrt)	24	Writing (Wrt)	25	Writing (Wrt)	27	Writing (Wrt)	31
Games (Gms)	25	Search (Srch)	25	Search (Srch)	25	Search (Srch)	27
Browsers (Brw)	24	Games (Gms)	23	Business (Bsn)	22	Browsers (Brw)	24
Search (Srch)	22	Browsers (Brw)	20	Games (Gms)	22	Games (Gms)	23
Business (Bsn)	18	Business (Bsn)	20	Browsers (Brw)	21	Business (Bsn)	23
Instant Messaging (IM)	11	Instant Messaging (IM)	12	Instant Messaging (IM)	13	Instant Messaging (IM)	14
Video (Vid)	10	Video (Vid)	11	Video (Vid)	12	Calendar (Cal)	13

TABLE 4. Top 10 sequences in 30-minute window data for negative mood.

[N_{bound}=1, N=191]		[N_{bound}=2, N=158]		[N_{bound}=3, N=130]		[N_{bound}=4, N=106]	
sequence	support (%)	sequence	support (%)	sequence	support (%)	sequence	support (%)
(Srch),(Srch)	20	**(Srch),(Srch)**	25	**(Srch),(Srch)**	27	**(Srch),(Srch)**	31
(Ref),(Ref)	19	(Ref),(Ref)	23	(Ref),(Ref)	27	(Ref),(Ref)	31
(Srch,Ref)	17	**(Srch,Ref)**	20	**(Ref,Srch)**	24	**(Ref,Srch)**	26
(SN),(SN)	16	(SN),(SN)	20	(SN),(SN)	22	(SN),(SN)	26
(Srch),(Ref)	15	**(Srch),(Ref)**	18	**(Srch),(Ref)**	21	**(Srch),(Ref)**	24
(Ref),(Srch)	14	**(Ref),(Srch)**	17	**(Ref),(Srch)**	20	**(Ref),(Srch)**	24
(Bsn),(Bsn)	13	(Bsn),(Bsn)	15	(SN),(SN),(SN)	19	**(Ref),(SN)**	22
(SN),(SN),(SN)	13	(SN),(SN),(SN)	15	**(Ref),(SN)**	18	(SN),(SN),(SN)	22
(Ref),(SN)	12	**(Ref),(SN)**	15	(Ref),(Ref),(Ref)	18	(Ref),(Ref),(Ref)	22
(SN),(Ref)	12	(Ref),(Ref),(Ref)	15	**(SN),(Ref)**	17	(Bsn),(Bsn)	21

quite high support values for 'Email' (Em) and 'Reference & Learning' (Ref). The 'Instant Messaging' (IM) and 'Video' (Vid) category only exist in the top ten of positive mood, whereas 'News & Opinion' (Nw) and 'Entertainment' (Ent) are only present in the top negative mood list. The other top categories, such as 'Business', 'Writing', 'Games' and 'Browsers', are present in both negative and positive mood with relatively smaller support values compared to the top three categories.

Sequence pattern results with minimum length 2-sequence are presented in Table 4 and Table 5. There are noticeable differences between top sequences in negative and positive mood labels. Negative mood has *<(Srch), (Srch)>* sequence with the highest support value whereas this sequence is not present in the top ten positive mood patterns. This is caused by the fact that 'Search' is much more common in negative mood patterns. Also, 'Games' (Gms) sequences are only present in positive mood patterns. Interestingly, sequences with 'Email' (Em) category only exist in the positive mood despite the fact that this category is common in both mood labels (support values higher than 25%). Similarly, 'Writing' (Wrt) sequences (e.g., *<(Wrt),(Wrt>*) are not present in negative mood patterns even though the category support values are relatively similar in positive and negative moods.

We also note some pattern similarities between these two mood labels. Two categories, such as 'Social Networking' (SN) and 'Reference & Learning' (Ref), have sequences occurring often in both positive and negative mood patterns. For example, sequence *<(SN),(SN)>* and *<(Ref),(Ref)>* are present in both label with relatively high support values. However, mixed sequences of these two categories, such as *<(SN),(Ref)>* or *<(Ref),(SN)>*, only occurred often in the negative ones. Also, the (SN) category formed a sequence with 'Email' (Em) in positive mood whereas in negative mood sequence (Ref) formed a sequence with 'Search', implying that the context in which they occur may have a link with mood. Beside forming with (SN), (Em) also formed with other digital categories in positive mood including *<(Em),(Ref)>*, *<(Ref),(Em)>* and *<(Em),(Wrt)>*. From the table, we can also find patterns with 3-sequence length, such as *<(SN),(SN),(SN)>* and *<(Ref),(Ref),(Ref)>*. These patterns are popular in both moods.

We also ran the algorithm using 60-minute, 90-minute and 120-minute time windows (not shown), the results are fairly similar to the 30-minute results. The main difference is that the support values are higher as the window size increases. Further, several categories, such as 'Games', ranked lower because the duration of this category became proportionally smaller compared to other categories.

Our results show that digital sequences could reveal information that would be hidden in duration-only analysis. For example, our data analysis shows that both email and search have relatively similar duration on both positive and negative mood labels. However, we found that there are many Email (Em) related frequent patterns in positive mood while Search (Srch) related patterns were dominant in negative mood. Therefore, finding frequent digital patterns can help

TABLE 5. Top 10 sequences in 30-minute window data for positive mood.

[N_{bound}=1, N=202] sequence	support (%)	[N_{bound}=2, N=171] sequence	support (%)	[N_{bound}=3, N=143] sequence	support (%)	[N_{bound}=4, N=121] sequence	support (%)
(Ref),(Ref)	23	(Ref),(Ref)	27	(Ref),(Ref)	31	(Ref),(Ref)	33
(SN),(SN)	22	(SN),(SN)	26	(SN),(SN)	29	(SN),(SN)	29
(Em),(Em)	19	**(Em),(Em)**	22	**(Em),(Em)**	25	(Ref),(Ref),(Ref)	26
(Em,Ref)	17	(Ref),(Ref),(Ref)	19	(Ref),(Ref),(Ref)	23	**(Em),(Em)**	25
(Ref),(Ref),(Ref)	16	**(Em,Ref)**	19	**(Em,Ref)**	21	**(Ref,Em)**	22
(Wrt),(Wrt)	14	**(Wrt),(Wrt)**	17	**(Wrt),(Wrt)**	19	**(Wrt),(Wrt)**	22
(Gms),(Gms)	14	**(Gms),(Gms)**	17	**(Em),(Ref)**	19	**(Em),(SN)**	20
(Em),(Ref)	14	**(Em),(Ref)**	17	**(Em),(SN)**	18	**(Em),(Ref)**	19
(Em),(SN)	13	**(Em),(SN)**	15	**(Ref),(Em)**	17	**(Ref),(Em)**	18
(Ref),(Em)	13	**(Ref),(Em)**	15	(SN),(SN),(SN)	17	**(Em),(Wrt)**	18

TABLE 6. Performances of various mood detection models.

Model	Classifier	Acc.	TPR	TNR	Prec.	F1
Overall self-reports proportion	-	56.8%	-	-	-	-
Baseline	ZeroR	71.1%	100%	0%	71.1%	83.1%
Raw duration (RD) [12]	Random Forest	66.7%	84.4%	23.1%	73.0%	78.3%
	Gradient Boosting	64.4%	78.1%	30.8%	73.5%	75.7%
Duration percentage (DP) [20]	Random Forest	68.9%	75.0%	53.8%	80.0%	77.4%
	Gradient Boosting	68.9%	75.0%	53.8%	80.0%	77.4%
Digital sequence patterns (SP)	Random Forest	71.1%	78.1%	53.8%	80.6%	79.3%
	Gradient Boosting	66.7%	68.8%	61.5%	81.5%	74.6%
Combination of duration and sequence (DP+SP)	Random Forest	80.0%	90.6%	53.8%	82.9%	86.6%
	Gradient Boosting	64.4%	68.8%	53.8%	78.6%	73.4%

provide a different angle on understanding the relationship between digital footprints and wellbeing.

B. MOOD DETECTION MODEL

After training all models using the two types of classifiers (Random Forest and Gradient Boosting) on 60-minute window data with $N_{bound} = 6$ and the corresponding self-reported mood labels, we retrieved the highest accuracy model from the Random Forest classifier for all models as shown in Table 6. Random Forest is known to be robust to outliers and non-linear data as it is based on decision trees. The classifier is also well-suited for high dimensional data and handles unbalanced data well. To evaluate performance for these models, we show for baseline the overall self-reports proportion for the whole self-reports and the ZeroR classifier performance. The overall proportion shows that 56.8% of self-reports were labelled as positive class. It indicates that positive mood label/class were more dominant compared to negative mood class. The baseline (ZeroR classifier) shows that the positive class is much more dominant in the test case. This baseline classifier always predicts the mood as positive (the most dominant in the training dataset). The classifier has an accuracy of 71.1% with 100% TPR as it did not miss any positive class. However, the TNR value is 0% because it missed all negative class.

Let us first examine the results for the RD and DP models. From Table 6, the RD model shows a maximum accuracy of 66.7% from Random Forest classifier. While the accuracy of this model is poor, the TPR value is still high (84.4%) as the model rarely missed the positive class. Similarly to the baseline classifier, this model struggled to predict the negative mood class, reaching only 23.1% TNR. Comparing to the RD model with Gradient Boosting classifier with the same features, the TNR and precision performance was better (30.8% and 73.5% respectively), but it has a lower TPR score. This means that the model tends to classify the test dataset into the negative class compared to the Random Forest classifier.

The DP mood prediction model performances show no difference between Random Forest and Gradient Boosting classifiers. The accuracy of these models (68.9%) is higher than the RD models (using features from [12]). The DP models have much higher TNR score 53.8%, but they struggled in predicting positive class with 75% TPR score (lower than the previous models). It is also important to note that these models have a better precision score of 80%. Overall, our results show that all models with features from [20] could not yield a higher accuracy than the ZeroR baseline.

The most prominent frequent digital patterns were selected as machine learning features based on the previous GSP results, and their impact on accuracy measured. The features were grouped into four types as shown in Table 7. Features are all binary (1 if the sequence(s) is present in the particular activity record, else 0). For example, the following activity record *<(SN, Ref), (SN, Em)>* will generate '1' for *<(SN),*

TABLE 7. Selected features for mood detection model.

Type	Feature
Top categories	'Search' / (Srch) 'Social Networking' / (SN) 'Email' / (Em) 'Games' / (Gms) 'News & Opinion' / (News) 'Writing' / (Wrt)
Top frequent negative mood patterns	(Srch),(Srch) (Srch),(Ref) (Ref),(Srch) (Ref),(SN) (SN),(Ref)
Top frequent positive mood patterns	(Em),(Em) (Gms),(Gms) (Em),(SN) (Em),(Ref) (Wrt),(Wrt)
Top frequent in both moods	(SN),(SN) (Ref),(Ref)

(SN)> and *<(Ref), (SN)>* features, but '0' for the *<(Em), (Em)>* feature.

Results show that the SP model with digital pattern features could reach an accuracy of 71.1% with Random Forest classifier. This accuracy performance is higher than the RD and DP models. Even though the accuracy is similar to the baseline classifier (ZeroR), the model has much better TNR score (53.8%) and precision score (80.6%) compared to the baseline. The same model with Gradient Boosting classifier yielded poor accuracy (66.7%). However, this particular model has the best TNR score (61.5%) indicating better prediction on negative class. The model also has a good precision score when predicting positive class (81.5%).

Lastly, the DP+SP model, which combines both digital patterns (SP) and duration percentage (DP) features reached the highest accuracy of 80% from Random Forest classifier. This model also has the highest TPR score (90.6%), the highest precision score (82.9%) and the highest F1 score (86.6%) compared to other models. The TNR score, however, (53.8%) is similar to other previous models (DP and SP models). This result suggests that both types of features (duration and patterns) are important and that their combination improves performance of mood detection.

VI. DISCUSSION

Some of the results described confirm past studies and others provide new insights. For example, total usage shows that social networking is correlated with more positive than negative mood reports. This result confirms those of [8] that suggested that Facebook might be a "light" interaction experience that acts as a "break" after being highly engaged with work.

Also, the games category has high support values in the positive mood label, a result similar to previous work investigating the use of computer games during working hours [6]. The results also suggest that negative mood reports are associated with frequent use of the 'Search' category, including activities like Google and Bing Search, etc.

Beyond replicating earlier studies, our sequence pattern methodology produced significant new findings. One of them is that the email-including sequence patterns (e.g., (Em),(Em) and (Em),(Ref)) only occurred in positive mood despite that email is common in both mood labels. A past study showed that stress can be reduced by checking less frequently [4], but our results show that people with positive mood have email-related sequences with high support values. This result indicates that people with positive mood usually use email consistently. However, we must note that Kushlev and Dunn's study asked their participants to answer stress-related questionnaire every 5PM on each weekday. In our study, we ask them randomly during the day and use shorter time window (30, 60, 90, and 120 minutes) whereas the other study used a working hour time window (around 8-9 hours). The other finding is that Search (Srch) related patterns (e.g., (Srch),(Srch) and (Srch),(Ref)) were more dominant in negative mood despite that the positive mood has a high support value on Srch category. These findings support our first hypothesis that sequence patterns can reveal patterns that are obscure from duration-only analysis. We can also find patterns that are more frequent prior to positive moods than prior to negative moods, and vice-versa.

Confirming our second hypothesis that digital sequence patterns are good predictors to detect mood, our mood detection results demonstrate that digital patterns can be used as features for mood detection systems. Compared to past studies, our model produced better accuracy results while using a general model (combining data from all participants). The model using features from [12] (RD) yielded a significantly lower accuracy for general model with digital activity duration data as the features. We acknowledge that the differences between the datasets might cause such effect. Our DP+SP model, which combines digital patterns and updated duration features, could reach an accuracy of 80%. This result is also higher than the model built using past study features (DP) [20] as shown in Table 6. This paper [20] shows an accuracy of 82-83%, but the results were from dataset with $N_{bound} = 6$ for 30-minute window and $N_{bound} = 11$ for 60-minute window. Using higher N_{bound} sizes would increase the performance for all models. However, we used datasets from $N_{bound} = 6$ from 60-minute window (at least 50% data are available) in this study. These datasets are more likely where missing digital data is common. Regardless, the increase of accuracy in combined features (DP+SP) model shows that digital pattern features are important to help predict mood. In the end, the method proposed in this paper can be applied for future research that analyses digital activity records, especially the ones that use RescueTime data. Also, this promising result on predicting mood suggests that these digital patterns can also be applied for other wellbeing related predictors, such as stress or mental health detection.

For performance testing (shown in Figure 2), data on the previous digital patterns were used in a model that predicted moods. After experimenting with another splitting approach based on the users (different users for training and testing), we found that digital patterns were distinctive and unique for each person while the duration data were more alike. The model would perform worse for new users with different digital patterns, but the model will have better accuracy as the model learns about the new users' patterns. Nevertheless, a model that combined both duration and digital pattern yielded a slightly better performance compared to the model with duration features only. Further, we only selected 12 digital patterns as features. This arbitrary number is selected based on the table shown in Table 4 and 5. Adding more digital patterns can improve the model's performance. Future studies might consider to explore more on how to find the best digital pattern number that will produce the best accuracy. Another aspect to explore is to include the number of occurrence for each pattern.

Both digital sequence pattern mining results and mood detection model can be improved in several ways. One improvement is to get more datasets for both GSP inputs and mood detection model. While having 1,244 mood self-reports from 72 participants showed promising results, we expect better accuracy by collecting more data from more participants. Future studies could also try to different features from both duration information and frequent digital activity patterns to build the mood detection model.

Another improvement is to capture more information, such as the context of each digital category use. For example, people can use social media for recreational, communicating with other people, or even for work if they work as sales or similar roles. Non-digital activities were not considered and investigated in this paper.

This study used a simple mood widget instead of PANAS [28] or other mood/emotion-related questionnaire since we did not give any rewards to our participants, the interface had to be kept simple and motivating. Finally, our study merged all data from all participants instead of analysing it per participant. This has been a common approach in past studies to see general digital activity impacts on wellbeing.

Ground truth data for the mood classification came from self-reports with five mood labels. Subsequently merging them into two classes (negative and positive) likely minimised the effect of potential errors in the data collection. Classification with more than two labels would be more sensitive to inaccurate mood reports. At this scale, collecting data from self-annotation is the most applicable approach to collect this kind of data, but future studies may consider different ways.

Our work explored various activity records window and N_{bound} sizes. We do not claim any optimal size for activity record time window and N_{bound} but we explored what information can be gathered and what results would be obtained by using our proposed preprocessing with various activity record time window and N_{bound} sizes. The process relies on human judgement and the analysis can provide different insights. Future research may choose to use different sizes to get a different analysis.

The aim of the paper has been to explore novel ways to decompose footprint data that occur just before mood self-reports and investigate potential relationships between digital activity sequences and mood. Cost and efficiency were not an issue at this stage of the research. The findings show that particular patterns were associated (high support value) with specific moods. Whilst these results do not prove nor suggest a causality between patterns and mood, they indicate a relationship that can serve as a basis for future work.

Moods are a key aspect of *hedonic* wellbeing, the focus of this study. Future work will explore *eudaimonic* wellbeing, using psychological models that consider different 'spheres of experience' (e.g., [29]), and how time-frames of experiences, relate to users' engagement with technologies. This approach provides a complementary perspective, clearly differentiating tasks from goal-directed behaviours. A related approach used in Human-Computer Interaction research is activity theory [30].

Methods that can be used to automatically infer states of wellbeing are important, but raise ethical challenges. On the one hand, understanding the impact of digital behaviour patterns is important to inform design. By knowing the association between particular digital patterns and wellbeing, we can design technologies that support rather than hinder wellbeing [1], [29]. This can be done, for example, in a preventative way, designing to make those behaviours less likely. But the use of behavioural data to predict moods also raises ethical challenges. While in our study all participants where informed of how their data would be used, this is not yet common in industry applications. We believe that for this research to have positive outcomes, users should always be informed that their behaviour is being monitored, and the purpose (e.g., research). This will allow users to judge if the technology offers a fair value, for example, by learning how their behaviours impact their wellbeing.

ACKNOWLEDGMENT

We thank Ehssan Sakhaee for his valuable comments.

REFERENCES

[1] R. A. Calvo and D. Peters, *Positive Computing: Technology for Wellbeing and Human Potential*. Cambridge, MA, USA: The MIT Press, 2014.

[2] G. R. Oldham, A. Cummings, L. J. Mischel, J. M. Schmidtke, and J. Zhou, "Listen while you work? Quasi-experimental relations between personal-stereo headset use and employee work responses," *J. Appl. Psychol.*, vol. 80, no. 5, pp. 547–564, 1995.

[3] J. Waycott, H. Davis, F. Vetere, A. Morgans, A. Gruner, E. Ozanne, and L. Kulik, "Captioned photographs in psychosocial aged care: Relationship building and boundary work," in *Proc. 32nd Annu. ACM Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 2014, pp. 4167–4176.

[4] K. Kushlev and E. W. Dunn, "Checking email less frequently reduces stress," *Comput. Hum. Behav.*, vol. 43, pp. 220–228, Feb. 2015.

[5] K. B. Wright, B. Abendschein, K. Wombacher, M. O'Connor, M. Hoffman, M. Dempsey, C. Krull, A. Dewes, and A. Shelton, "Work-related communication technology use outside of regular work hours and work life conflict: The influence of communication technologies on perceived work life conflict, burnout, job satisfaction, and turnover intentions," *Manage. Commun. Quart.*, vol. 28, no. 4, pp. 507–530, 2014.

[6] L. Reinecke, "Games at work: The recreational use of computer games during working hours," *CyberPsychol. Behav.*, vol. 12, no. 4, pp. 461–465, 2009.

[7] E. K. Yuen, E. A. Koterba, M. J. Stasio, R. B. Patrick, C. Gangi, P. Ash, K. Barakat, V. Greene, W. Hamilton, and B. Mansour, "The effects of Facebook on mood in emerging adults," *Psychol. Popular Media Culture*, vol. 8, no. 3, pp. 198–206, 2018.

[8] G. Mark, S. Iqbal, M. Czerwinski, and P. Johns, "Capturing the mood: Facebook and face-to-face encounters in the workplace," in *Proc. 17th ACM Conf. Comput. Supported Cooperat. Work Social Comput. (CSCW)*, New York, NY, USA, 2014, pp. 1082–1094.

[9] A. M. Shaw, K. R. Timpano, T. B. Tran, and J. Joormann, "Correlates of Facebook usage patterns: The relationship between passive Facebook use, social anxiety symptoms, and brooding," *Comput. Hum. Behav.*, vol. 48, pp. 575–580, Jul. 2015.

[10] E. I. M. Collins, A. L. Cox, J. Bird, and C. Cornish-Tresstail, "Barriers to engagement with a personal informatics productivity tool," in *Proc. ACM 26th Austral. Comput.-Hum. Interact. Conf. Designing Futures, Future Design (OzCHI)*, New York, NY, USA, 2014, pp. 370–379.

[11] G. Mark, Y. D. I. Wang, M. Niiya, and S. Reich, "Sleep debt in student life: Online attention focus, Facebook, and mood," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst. (CHI)*, New York, NY, USA, 2016, pp. 5517–5528.

[12] R. Ferdous, V. Osmani, and O. Mayora, "Smartphone app usage as a predictor of perceived stress levels at workplace," in *Proc. 9th Int. Conf. Pervasive Comput. Technol. Healthcare (PervasiveHealth)*. Brussels, Belgium: ICST, 2015, pp. 225–228.

[13] G. Mark, M. Czerwinski, S. Iqbal, and P. Johns, "Workplace indicators of mood: Behavioral and cognitive correlates of mood among information workers," in *Proc. ACM 6th Int. Conf. Digit. Health Conf. (DH)*, New York, NY, USA, 2016, pp. 29–36.

[14] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5 (R))*. Philadelphia, PA, USA: American Psychiatric Association Publishing, 2013.

[15] D. Kahneman and A. B. Krueger, "Developments in the measurement of subjective well-being," *J. Econ. Perspect.*, vol. 20, no. 1, pp. 3–24, 2006.

[16] P. Paredes, D. Sun, and J. Canny, "Sensor-less sensing for affective computing and stress management technology," in *Proc. IEEE 3rd Int. Workshop Pervasive Comput. Paradigms Mental Health*, May 2013, pp. 459–463.

[17] M. Ciman and K. Wac, "Individuals' stress assessment using human-smartphone interaction analysis," *IEEE Trans. Affective Comput.*, vol. 9, no. 1, pp. 51–65, Jan./Mar. 2018.

[18] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," in *Proc. 5th Int. Conf. Extending Database Technol., Adv. Database Technol. (EDBT)*, London, U.K.: Springer-Verlag, 1996, pp. 3–17. [Online]. Available: http://dl.acm.org/citation.cfm?id=645337.650382

[19] D. Kahneman, E. Diener, and N. Schwarz, *Well-Being: Foundations of Hedonic Psychology*. New York, NY, USA: Russell Sage Foundation, 1999.

[20] M. J. Alibasa and R. A. Calvo, "Supporting mood introspection from digital footprints," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 1–6.

[21] T. Grover and G. Mark, "Digital footprints: Predicting personality from temporal patterns of technology use," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., ACM Int. Symp. Wearable Comput. (UbiComp)*, New York, NY, USA, 2017, pp. 41–44.

[22] M. S. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection," *Behav. Res. Methods, Instrum., Comput.*, vol. 32, no. 1, pp. 93–110, Mar. 2000.

[23] O. Brdiczka, N. M. Su, and B. Begole, "Using temporal patterns (t-patterns) to derive stress factors of routine tasks," in *Proc. ACM CHI Extended Abstracts Hum. Factors Comput. Syst. (CHI EA)*, New York, NY, USA, 2009, pp. 4081–4086.

[24] A. M. Isen, M. Clark, and M. F. Schwartz, "Duration of the effect of good mood on helping: 'Footprints on the sands of time,'" *J. Personality Social Psychol.*, vol. 34, no. 3, pp. 385–393, 1976.

[25] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaïane, "Clustering and sequential pattern mining of online collaborative learning data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 6, pp. 759–772, Jun. 2009.

[26] P. Fournier-Viger, A. Gomariz, T. Gueniche, A. Soltani, C.-W. Wu, and V. S. Tseng, "SPMF: A java open-source pattern mining library," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3389–3393, Jan. 2014. [Online]. Available: http://dl.acm.org/citation.cfm?id=2627435.2750353

[27] RescueTime. *How is My Productivity Pulse Calculated?* Accessed: May 26, 2019. [Online]. Available: https://help.rescuetime.com/article/73-how-is-my-productivity-pulse-calculated

[28] V. Tran, "Positive affect negative affect scale (PANAS)," in *Encyclopedia of Behavioral Medicine*. New York, NY, USA: Springer, 2013, pp. 1508–1509.

[29] D. Peters, R. A. Calvo, and R. M. Ryan, "Designing for motivation, engagement and wellbeing in digital experience," *Frontiers Psychol.*, vol. 9, p. 797, May 2018.

[30] V. Kaptelinin and B. A. Nardi, *Acting With Technology: Activity Theory and Interaction Design*. Cambridge, MA, USA: MIT Press, 2006.

MUHAMMAD J. ALIBASA received the B.Sc. and M.Sc. degrees in electrical engineering from the Institut Teknologi Bandung, Indonesia, in 2014 and 2012, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science, The University of Sydney, Australia. He was a Mobile Application Developer and developed mobile apps in the area of wellbeing. His research interests include human-computer interaction and artificial intelligence. He was a recipient of the Indonesian Endowment Fund for Education (LPDP) scholarship from Ministry of Finance, Republic of Indonesia.

RAFAEL A. CALVO is currently a Professor with the Imperial College London and The University of Sydney. He is also the Director of the Wellbeing Technology Lab and the Software Engineering Group that focuses on the design of systems that support wellbeing in areas of mental health, medicine, and education. He is a keen player in the push for more ethical technology design. In 2019, he will join the Leverhulme Centre for the Future of Intelligence. In 2015, he was appointed as a Future Fellow of the Australian Research Council to study the design of wellbeing-supportive technology. He is a member of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, a part of the UN's AI for Good Global Summit.

KALINA YACEF received the B.Sc. degree in maths and the Ph.D. degree in computer science. She is currently an Associate Professor with the School of Computer Science, The University of Sydney, Australia. She leads the Human-Centered Data Mining for Education and Health & Wellbeing. She serves on the Board of the International Educational Data Mining Society (IEDMS) and the Executive Committee of the International Artificial Intelligence in Education Society (IAIED). She is an Associate Editor of the IEEE Transactions on Learning Technologies and *International Journal of Educational Data Mining*.