

The Application of User Event Log Data for Mental Health and Wellbeing Analysis

Robin Turkington
Ulster University Jordanstown
Shore Road, Newtownabbey
Northern Ireland
turkington-r@ulster.ac.uk

Maurice Mulvenna
Ulster University Jordanstown
Shore Road, Newtownabbey
Northern Ireland
md.mulvenna@ulster.ac.uk

Raymond Bond
Ulster University Jordanstown
Shore Road, Newtownabbey
Northern Ireland
rb.bond@ulster.ac.uk

Siobhan O'Neill
Ulster University Coleraine
Cromore Road, Coleraine
Northern Ireland
sm.oneill@ulster.ac.uk

Cherie Armour
Ulster University Coleraine
Cromore Road, Coleraine
Northern Ireland
c.armour1@ulster.ac.uk

Many digital interaction technologies, including web-based interventions, smartphone applications, and telephone helplines, can provide a basis for capturing real time data of interactions between the user and the system. Such data is recorded in the form of log files, which records user events that range from simple keystrokes on a computer, user activated sensor data or duration/frequency of phone calls. These interactions can provide rich datasets amenable to user data analytics using machine learning and other analytics techniques. This data analysis can highlight usage patterns and user behaviours based on their interaction with the technology. User log data analysis can be descriptive statistics (what users have done), predictive analytics (what events will happen) and prescriptive (what action to take given a predicted event or outcome). This can also be thought of as spanning across different levels of user analytics from hindsight, insight and foresight. Predictive analytics are used with log data to provide predictions on future user behaviour based on early usage behaviours. Event logs are objective regarding usage, but usage may not correlate with the level of the system's user experience. Hence, ecological momentary assessment (EMA) of the user experience can be used augment user log data. Nevertheless, with the emergence of health applications and other app-based health services, we consider how user event logs can be specifically used within the mental health domain. This can provide beneficial insights into how users interact with mental health e-services, which can provide an indication of their current and future mental state.

User event log, log analytics, machine learning, data mining, mental health, digital interaction technology, m-Health, e-Health

1. INTRODUCTION

1.1 Description of event logs

Event logs can be described as a digital record of a system's alerts and notifications. Event logs are supported by many applications, call services, network devices and other technology-based systems and services. This provides a rich stream of data for analysis by stakeholders. The bulk of event log analysis research has looked at security applications

(Vaarandi, 2006; Vaarandi & Pihelgas, 2014; Vaarandi & Grimaila, 2012), health care (Mans et al 2008, 2014) and business (Armas-Cervantes et al, 2017; Tax et al, 2017) which have generally been used to capture and 'debug' incidences that occur within an electronic system. In a human-computer interaction capacity, an "event" can be described as an occurrence within a system of any interaction between a human and a digital device or call line such as a computer, a smartphone application, wearable device or call line service (Olson & Kellogg, 2014). For

this paper, we refer to event logs as interactions between an individual and a digital service. Indeed, a single event log can have many different purposes. A log file can provide information about user traffic on a network path, and these can provide insight to help optimise performance of a service, or they can also be used to form logs, identifying who and when a person calls a service, call volume, reason, duration and location (Oinas-Kukkonen, 2013).

Event log data is initially collected at device level and may be transferred at an aggregate level to a remote destination, e.g., in the 'cloud'. Each log file entry is assigned a unique sequential log entry number that acts as an identification key for that record. A log entry can also be thought of as a "stamp", representing an action made by the user of the system or by the system itself. An event log must contain these fields: a unique identifier code which may be anonymised; a field for the event that was recorded; and a data and time stamp field. Normally this is at the discretion of the system designer and may be configurable by the developer/researcher to contain other information fields.

1.2 What can a user event log tell us about a user of a service?

Event log data is generated instantly as the user interacts with the system. These events are recorded in real time, breaking down how the user interacts with the system. A user session can last for seconds, minutes, hours or even days. When event logs are analysed, researchers may wish to select any number of features as metrics that matter to the user or the service stakeholders (Olsen & Kellogg, 2014). These user metrics may be a descriptive overview on user activity such as the frequency of activities, i.e. frequency of logins, the most and least popular feature(s) of the system, activities started/completed etc. From this information, the researchers and developers can determine the extent to which the system resonates with the user, what features of the system is the most persuasive and what features of the system requires improvement. Other metrics that are not explicitly available within the raw data can also be analysed. For instance, the Lumiere project (Horvitz et al, 1998) explored the effectiveness of probabilistic techniques to assess user interactions with Microsoft Office applications. The authors used an event monitoring system to capture a wide range of user actions, to which Bayesian methods were employed to predict the goals of the user by assessing their

background, the interactions between the user and the system and the queries in which the user made (Fox et al, 2005). Other studies have predicted user's interest levels through user reading duration (Morita & Shinoda, 1994; Konstan et al, 1997), and predict user's explicit satisfactory ratings of web search results through user implicit interactions (Fox et al, 2005).

1.3 The relationship between the user and the service

There are many reasons why an individual may use a service, but there are factors that may determine whether or not the service provides the optimum outcome; one factor of which is usability. Usability is an important mediator within many digital interventions, e.g., health applications, as it is important for services to be easy to use. An individual may consider a service to be usable based on whether or not the service is easy to understand, if the user is able to learn from the service and if the user is able to operate the service with ease (Zapata et al, 2015). Although user event logs can provide valuable insight into how the user interacts with the service, user event log data often assumes a dose-response relationship and mainly provides quantitative statistics. It is often deemed that the higher the frequency of specific interactions between the user and features of the service, the more popular that feature is to the service user; this could also be deemed as suboptimal use of the service itself. It is difficult to infer from the log data alone whether or not a frequent or infrequent user of the service finds it user friendly or unfriendly. If a person leaves a page without clicking on any features, this could mean either the individual could not find what they were looking for, or the features that were presented on the page satisfied the user's needs (Van Germert-Pijnen, Kelders & Bohlmeijer, 2014). The dilemma is whether user event log analysis reliably correlates with usability of the service. The following scenarios are possible:

- A frequent user who finds the service user friendly
- A frequent user who does not find the service user friendly
- An infrequent user who finds the service user friendly
- An infrequent user who does not find the service user friendly

While log data analysis is purely quantitative and does not provide the important qualitative knowledge of the participant, this can be

addressed using Ecological Momentary Assessment (EMA). In an electronic system, EMA involves the provision of prompts for the user in the form of questions, which can be used to gather user opinion or state in qualitative as well as quantitative form. Without EMA, such insights are not as obvious from examining the log data alone. In other kinds of studies that may involve log data as a means of data collection (i.e., laboratory test, diary, clinician rating), a system may gather opinions from the participant, but it may not be in real time or could be subject to recall bias. EMA complements user event log data as it can elicit user opinions and provide additional details that can explain user behaviours (Moskowitz & Young, 2006; Bellman, 2015).

2. INFORMATION DISCOVERY FROM EVENT LOGS

2.1 Data mining process

Event logs can provide researchers and stakeholders with rich streams of user data, to which this data can then be subjected to a structured data mining procedure to provide meaningful insights. Data mining is the extraction of information of interest, such as patterns or knowledge from huge amounts of data. There are many data mining techniques which can be utilised by analysts to conduct knowledge discovery tasks within event logs. An alternative term that describes data mining is Knowledge Discovery (mining) in Databases, or KDD. The generic process for KDD is as follows (Han, Pei & Kamber, 2011):-

1. Data cleaning: This is the process of removing noise and inconsistent data within the data set. This may involve removing outliers and accounting for missing data which could potentially distort the analysis outcomes.
2. Data integration: This is where many data sources of interest can be combined.
3. Data selection: The data analyst selects the relevant data for analysis from the data set.
4. Data transformation: The analyst then transforms the data into the appropriate form for mining.
5. Data mining: Machine learning methods are then applied to extract patterns within the data
6. Interpreting the discovered patterns: The patterns that represent knowledge that are of interest are analysed.

7. Knowledge presentation: This includes visual and knowledge representation techniques applied to the results to help explore patterns and trends within the data.

This data mining process can be applied to user event log data analytics. This has been referred to as Health Interaction Log Data Analysis (HILDA) which is similar to the Cross Industry Standard Process for Data Mining (CRISP-DM; Wirth & Hipp, 2000) and Knowledge for Discovery in Databases (KDD; Fayyad, Piatetsky-Shapiro & Smyth, 1996). Mulvenna and colleagues (Mulvenna et al, 2017) proposed H-ILDA, which has been designed to focus on data prospecting and includes machine learning phases for analysing interactional/event data from digital healthcare services. The H-ILDA process workflow includes data preparation, data prospecting and machine learning and is described in the following section.

2.2 H-ILDA process

In the data preparation phase, the log data is first subjected to a cleaning process, in which a data quality check is carried out to assess for missing data or any other anomalies, and removal of these if where possible. Once the data has been cleaned, it is then subjected to the next stage which involves the normalisation of the time stamps within the log data. The normalisation of the dates and times are relative to when the user started using the system, which allows for a temporal analysis of the event logs. This can be an important aspect particularly within health research as some illnesses are exacerbated as a result of the seasons (Mulvenna et al, 2017).

In the data prospecting stage, the log data's main characteristics are first summarised visually through histograms, probability density functions and correlations through bivariate analysis. The visualisation of this information can be regarded as means to explore or communicate with the data (Morrison & Doherty, 2014). Tukey (1977) encouraged exploratory data analysis through graphical plots to ensure understanding of the data before the application of relevant statistical techniques, and to understand particular trends and spot outliers within the data (Lam et al, 2007). Secondly, in this data prospecting stage relative time series analysis can be carried out. A time series can be regarded as the order of observations which is measured at different points in time. These measurements can be taken at any point under the discretion of the researcher, whether this is on the hour, day,

week, month or year (Montgomery, Jennings & Kulahci, 2015). The final machine learning stage may involve unsupervised and supervised machine learning processes. The aim of the unsupervised process is to model the underlying structure of the data to learn more about the user behaviours such as frequencies of actions. The log data is then subjected to a clustering analysis, with the outcome from the cluster analysis is reviewed through qualitative analysis by domain experts. Subsequently, feature extraction takes place from the initial user engagement data which is then used to predict future user types. This can allow researchers to identify at an early stage if the user is going to adhere to the service or stop using it after a short period of time.

2.3 What defines a “pattern” within event logs?

Discovering patterns within event logs has been regarded as an important process to uncover information about what patterns of user behaviour emerge when the users navigate through the service. Frequent patterns that are detected can reveal what types of events are closely related to each other, and this knowledge can be used for the creation of specific rules for event correlation methods (Vaarandi, 2005). A pattern mining algorithm may see the event log as a number of overlapping windows, where a window begins at a particular time X and contain events within a time frame of Y seconds (Vaarandi, 2005; Klemettinen, 1999). The resulting combination of events that occur within the time frame may be considered a frequent pattern if the combination is present in a number of windows. The number of windows required to determine the existence of a frequent pattern is decided by the researcher. The order of events in windows can provide the researcher with a direction of the relationship between event types and whether or not they are causal of one another. The patterns that emerge in usage of the service can be of value to understand the critical moments in which the employment of persuasive features of the intervention that support motivation, to predict if a user is going to adhere to the service or when there will be an increase/decrease of use of the service (Van Gemert-Pijnen et al, 2014).

There are various levels of insight that data analytics can produce from event log data. They can be used for the discovery of usage patterns which can be used to help predict future user behaviour. This has the potential for providing a more timely and bespoke

intervention, to encourage alternative behaviours or the reinforcement of behaviours.

Using the Gartner Analytics Ascendancy Model (Maoz, 2013), these different levels of insight can arguably fall under three different categories: descriptive, predictive and prescriptive analytics. It is reasonable to consider the extent to which user event log analytics can fall under each of these categories.

2.4 Descriptive analytics

Descriptive analytics refers to tasks that characterises different properties within datasets. These methods focus on deriving patterns that can summarise the underlying patterns within the event logs. Descriptive analytics is generally exploratory in nature. Some of the descriptive analytics techniques that can be applied to event logs are as follows:

2.4.1 Clustering

The aim of clustering is to automatically group data that are similar within a cluster, and group data are dissimilar in another group i.e. grouping similar users into respective clusters. Clustering techniques have been used across many disciplines due to their appeal and practicality in exploratory data analysis (Jain, Murty & Flynn, 1999). Clustering analysis has the potential to inform researchers and practitioners the users of the service that are more likely to become adherers or abstainers (Thickett, 2006 ;Mulvenna et al, 2017). These types of approaches assume that each event within a log is described by a single line within the event log, and each line pattern representing a group of similar events. Any lines that are detected by the clustering algorithm are considered as “infrequent” will form a cluster of outliers. Clustering helps to identify the types of users of the service within the data. The k-means method, which is a popular clustering technique, partitions N observations into clusters in where each observation belongs to a cluster which shares the nearest mean (Cui et al, 2014). However, some challenges remain in clustering analysis. Many event logs are highly dimensional meaning that data points within the event log can have many attributes. This has made it difficult for traditional clustering methods to group; an issue described by Richard Bellman as “the curse of dimensionality” (Bellman, 2015; Assent, 2012). Issues that may occur with high dimensionality resulting from a fixed number of data points will become increasingly sparse as the dimensionality increases and consequently the algorithm identifies clusters

within the dataset that are not meaningful. The issue with the increase in sparsity is that clustering algorithms depend on the measurement of distance between data points and that objects in one cluster are closer to each other than to other objects in different clusters (Steinbach, Ertöz & Kumar, 2004).

2.4.2 Association analysis

This refers to the discovering of patterns within the data that describe relationships. The uncovered relationships within the data set can be represented in the form of association rules which are used to discover what features of a digital health intervention are used together, or indeed which events occur together. Even if these features are not directly linked, it can reveal what type of features are most common among users. Developers can use association analysis to determine what features within digital health interventions, when used together, are persuasive/un-persuasive in changing behaviour or improving health outcomes (Van Gemert-Pijnen et al, 2014; Iváncsy & Vaik, 2006). An established algorithm for association rule mining is the apriori algorithm (Vaarandi and Pihelgas, 2015).

2.5 Predictive analytics

The objective of predictive analytics is to predict the value of an attribute that is based on the values of other attributes (also known as independent and dependent variables or predictors and target features). Taking attributes from previously mined data from event logs and using them as independent variables, it becomes possible to classify and predict future users within the system. There are many machine learning techniques that can be applied to user event data for predicting future behaviour of the user behaviour. For instance, Hidden Markov Models (HMM) are a ubiquitous tool for modelling time series data and are used as a technique for representing probability distributions over sequences of observations within a dataset (Oliner, Ganapathi & Xu, 2012). HMM's can provide health practitioners with a time to future events i.e. give an approximate time to when a patient may have a lapse/episode in their treatment process (Lam et al, 2007). The two main types of HMM techniques are Discrete-Time (DT-HMM), designed for analysing event data of a regular nature, while Continuous-Time (CT-HMM) is designed for analysing event data that occurs at irregular data (Dempsey et al, 2017). Regression analysis techniques can also be used to assign observations to a data set. Linear regression is a supervised machine learning technique in which it is used to predict

values from a continuous range, whereas logistics regression is a classification algorithm that can make observations from discrete data (Witten et al, 2016). Two of the most popular predictive methods that are applied to event log data are discussed below:

2.5.1 Support Vector Machines (SVM)

SVM's have applications in pattern recognition, prediction, regression analysis and feature selection, making SVM's a popular choice within research involving classification. Their popularity is due to their efficiency in classifying non-linearly separable data by using a "kernel function". The introduction of the kernel enables the SVM to become flexible in separating the data into different classes (Auria & Moro, 2008; Al Machot et al, 2017). The linear kernel function operates by creating a hyper-plane within a high dimensional space. This hyper-plane is then used to separate the data into classes in a way that it fits in the middle of the gap between the classes and is maximally far away from both classes of data (Miner et al, 2014). For non-linear data, the polynomial kernel function allows the modelling of feature conjunction are up to the order of polynomial, by classing the data into circles (or hyperspheres; Al Machot et al, 2017). SVM's are recognised as being promising in terms of accuracy for identifying and discriminating patterns of user activity within unseen event data (Krishnan & Cook, 2014; Incel, Kose & Ersoy, 2013; Minor, Doppa and Cook, 2015). One limitation in applying traditional SVM methods to event data is that they can only carry out binary-classification, which may be an issue if a researcher wishes to classify more than two classes.

2.5.2 Decision trees

The technique uses an analogy of a tree to visually represent all possible options and consequences based on each option, with the aim of aiding classification of new data based on a series of hierarchical rules. The tree itself is drawn upside-down, with the root of the tree representing the initial condition. From there, the tree iteratively splits into branches and at the end of each branch represents a decision (conceptually known as a leaf). The branch that does not split anymore represents the final decision/classification (Mohri, Romstamizadeh & Talwalkar, 2012).

There are other supervised machine learning techniques such as neural networks and k-Nearest Neighbour. In user event log analysis, supervised machine learning techniques can be used to predict future behaviour or events based on previous behaviour or events elicited

early in the user lifecycle. It can also be used to predict the type of user (adopter, abstainer etc.). In mental health, supervised machine learning can be used to predict severity of depression based on user interactions.

2.6 Prescriptive analytics

Prescriptive analytics advise and suggest different possible actions towards a solution. Prescriptive analytics can not only predict what will happen but also provide recommendations regarding the actions that should be taken. Prescriptive analytics can continually re-predict to improve prediction accuracy. Regarding event logs, the prescriptive analytics for event sequences consist of recommended actions that would lead to an optimal outcome based on the history of previous events (Wiesner & Pfeifer, 2014). Within prescriptive analytics, recommender systems are used to help people with decision making, which is common in many domains such as recommending what products to buy next or what movies to watch next, as an example. Another example of where recommender systems have also been utilised within healthcare settings, as a means for health care practitioners to carry out patient-orientated decision making through the provision of scientifically proven or generally accepted medical information. Health recommender systems rely on the personal health history of the patient, with the goal of the recommender system to prescribe medical advice that is most relevant to the development of the patient.

3. USING EVENT LOG DATA IN MENTAL HEALTH RESEARCH

3.1 Applications

Mental health research has embraced event log data as a means of providing insight into assessing mental health and wellbeing within a population and providing real time information on the behaviour of those who use digital health technologies. Many digital health technologies are designed to help aid behaviour change amongst users to improve health outcomes. Such technologies allow for the tracking of the symptoms of various mental illnesses in real-time. As well, event log data can provide many different advantages to mental health research. For instance, an issue with traditional mental health research is that studies which involve randomised control trials (RCT) only collect participant data at fixed time points, limiting the insight to be gained in observing the progress of the user in real time. With RCTs, it can be difficult to gain insights

as to what are the mediating factors of the digital interventions that contribute more (or even least) to promoting behaviour change, adherence and overall mental wellbeing (Sieverink et al, 2017). Event log data is captured in real time which allows for insight into the current behaviour of the user/patient and provide a basis for researchers to analyse mediating factors which cause the change in user/patient behaviour.

Using diaries to record data can provide qualitative information on the mediating factors which contribute to behaviour change and allows researchers to understand the user/patient's experience through discourse and in a non-statistical way (Harvey, 2011). However, diary entries are not always recorded in the moment, with data often being entered in a retrospective manner. Once more, diary entries can be falsifiable, masking the true user/patient experience and are more time consuming to complete and can be considered less-engaging to complete (Jimoh et al, 2018). In addressing this, event logs have the capability to capture data in a manner which is more reflective of the user experience, in real time and in a way which is more exacting. In some instances, it could be beneficial for future behavioural research to compliment diary studies with event log data, as this may provide some validation for the qualitative information which is contained within the diary entries, while at the same time, provide additional and in-depth insight into the user event log data, particularly within an m-Health setting (Bradway et al, 2018). These are just some of the advantages in which event logs have over other methods within behavioural research. the capability to show researchers the mediating factors, and the course in which these factors unfold, within digital interventions that lead to an improvement in health-related outcomes.

Event logs can provide validation for existing psychological models and theories which underpin mental illness. Many features within digital health interventions for mental health have a theoretical basis within behavioural change literature. For instance, many mHealth interventions are based on Cognitive Behavioural Therapy (CBT; Szigethy et al, 2018; Furukawa et al, 2018; Mantani et al, 2017). It is then possible for psychiatrists and therapists to look for patterns within the log data which depict how the user is navigating through the service. The patterns of features that have been used (features that correspond to the behavioural theory that has been applied to the intervention) can be viewed through the log data (Sieverink et al, 2017).

Another advantage is that user log data can be considered as far more reliable than traditional data collection methods such as questionnaires, diaries and interviews (Olson & Kellogg, 2014). The latter methods of data collection can be subject to recall bias or influenced by the presence of researchers, affecting the accuracy of the data collected. Studies that are carried out in a laboratory setting may elicit unwanted performance effects from participants if “natural” behavioural phenomenon is to be observed. Field studies which aim to capture human behaviour may do so in a more naturalistic environment, but in doing so outside of laboratory conditions, researchers may have less control over trial. Event logs capture data “in the wild”, capturing natural observations and interactions of the user in their own environment, uninfluenced by the external factors as mentioned with the previous methodologies and not to the detriment of the researchers through diminishing control (Olsen & Kellogg, 2014).

With the usage of event logs, this can also allow practitioners to track patient progress and symptomology, which allows better treatment outside of clinical appointments (Walsh et al, 2016). For instance, the Diagnostic and Statistical Manual (DSM) outlines the clinical criteria for all known mental illnesses, to which the criteria assess behaviour as a direct symptom of the mental illness itself. A practitioner can see from a user's event log from a smartphone application that decreased physical activity, irregular sleep patterns and lowered social interactions are signs of depression (American Psychiatric Association, 2014; Aung, Matthews and Choudhury, 2017). This was demonstrated in a study by Saeb et al (2015), in which the authors demonstrated significant correlations between the user's Patient Health Questionnaire scores (PHQ-9; Kroenke & Spencer, 2002) and geo-locational data captured by the user's smartphone; the metrics used were 24 hour circadian movement, normalised location entropy (i.e. consistency of the user's presence between routinely visited locations), location variance, phone use duration and frequency maximum (Saeb et al, 2015; Aung, Matthews and Choudhury, 2017).

Event logs can provide foresight into the future mental health and wellbeing status of the user, through the recognition of patterns of previous log entries which can be of added value to the development of models that promote behaviour change. Using machine learning techniques, it is possible to make predictions on whether the user will continue with the service (Sieverink et al, 2017). A key mediator

to improving mental health is user adherence to the intervention. The term “adherence” denotes the level of which the user's activity within the intervention parallels the pattern of activity that was intended by the intervention developers (Donkin et al, 2013). Event log data within digital interventions can provide researchers with progress updates in real time and provide an objective insight into how users adhere to the intervention for their needs. The number of logins, actions of the user and the amount of completed lessons/modules can be considered as a metric across a range of studies to identify any differences between users who are potential adherers/non-adherers to the intervention (Van Gemert-Pijnen et al, 2014; Kelders et al, 2016; Freyne et al, 2012; Kelders et al, 2013).

Event log research has also examined patterns in call log data within crisis helplines. These helplines assist callers who are experiencing psychological distress by providing emotional support and advice. A crisis is defined as a state of psychological disruption in where the individual's usual coping mechanisms are ineffective (Spittal et al, 2015; Kalafat et al, 2007). Call log data can help to profile the types of callers that use the system, as in whether they are prolific or infrequent callers, which is defined by the amount of calls a particular user makes to the service and mean duration of those calls. Factors that contribute to psychological crisis of the users can be recorded within call logs, which provides an insight into the reason(s) why someone may use the crisis service. The number of calls that a person in crisis makes to these helplines is indicative of help-seeking behaviour and can also provide a prediction on whether the user is likely to die by suicide. Characteristics that appear within the call log data in the group of users who have taken their own life could indicate risk factors for suicide in the population of service users (Bunting et al, 2015). This vital information can improve the delivery of support through crisis helplines, allowing service operators to better identify mediating risk factors within the user population and therefore provide appropriate advice or a timely intervention. In some instances, the log data can contain ward location from which the user is calling from, although this is at the discretion of the helpline service. This information can provide insight into areas that suffer from most hardship (i.e. poverty, crime), which is a key predictor of mental illness (Eibner et al, 2004).

3.2 Example application: “Mobilyze!”: an mHealth intervention to treat depression

The following example illustrates how user event log data can be utilised in gathering insights into a user's current health state, their journey through rehabilitation, an instance in how event log data can be applied to machine learning applications and how event log data validates existing scales used to assess behaviour change. Burns and colleagues (2011) examined how the user sensor data collected through a mHealth application "Mobilyze!" could be used to predict depressive symptoms in individuals. The application comprised of a context-aware system where through using machine learning algorithms (see Phase 2), was able to learn, predict and act on specific user mental states from at least 38 sensor values including the user's location, activity, social environment and internal states. Users were prompted to report their states using EMA on the phone application (i.e., mood, concentration, affect). The intervention was designed to detect when depressed patients may require assistance, through the analyses of the sensor values and internal state. The authors designed a "context-aware" system, consisting of three phases:-

- Phase 1: Phone sensors gathered observations about the participant and their environment. This included information such as user interaction data (i.e., number of calls made, applications used).
- Phase 2: Pruned Weka REPTrees (Hall et al, 2009), a fast decision tree learner, consisting of a decision/regression tree algorithm, was used to learn the relationship between the user sensor data and the user's reported social context, activity, location and internal states. Through EMA, the user was prompted to report their current mental state. When the intervention collected the EMA responses, it simultaneously recorded sensor to which the intervention paired with the EMA responses. From here, the machine learning algorithm created a participant specific model to predict the user's state in the future from the sensor data.
- Phase 3: The intervention application carried out an action phase, which relayed the prediction analysis to the participant; the authors stated that this action could also include the user's "coach".

The system required users to label their own states through selecting values for each context category, which was matched with the user's sensor data. The user may rate the

intensity of physical activity they are doing, their location, social context etc. The system would then match this information with the current sensor data. Each time a new state was entered by the user, a new model was generated. The intervention prompted the user to tell the system his/her state around 5 times per day. The users labelled their own states using a 7-point Likert scale to rate their overall mood, intensity of mood, pleasure, sense of accomplishment, cognitive state and perceived control; a 4-point Likert scale was used to measure physical exertion.

The Mini-International Neuropsychotic Interview (Hergueta, Baker & Dunbar, 1998) was used to evaluate the level of change in major depressive disorder diagnosis. The Quick Inventory of Depression Symptom-Clinician rating (QIDS-C; Rush et al, 2003) and Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer & Williams, 2001) were used to measure depression symptom severity; the Generalised Anxiety Disorder scale (GAD-7; Spitzer et al, 2006) was used as a secondary measure to assess for anxiety symptom severity. Outcomes were assessed at baseline, week 4 (mid-treatment) and week 8 (post-treatment).

The mean predictive accuracy for classification of location was 60.3%, but the other predictive values for user states (such as mood) were generally poor. The authors concluded that this may have been due to inaccuracies that are associated with the phone sensor data and that to be truly predictive, the raw sensor data has to be technically accurate and manipulated in order to extract meaningful data from it.

Participants within this study reported a significant improvement on self-reported and interview measures of depressive symptoms and were less likely to meet the criteria for major depressive disorder; anxiety symptoms also decreased over the course of treatment.

This study was one of the first to apply a new paradigm for a mHealth intervention in treating depression. While a context-aware system shows some promise, the authors concluded that their own system required further refinement and a need for continued development and further testing with this kind of technology.

4. CHALLENGES

4.1 The "Black Box" phenomenon

There is a growing need for closer collaboration between data science and cognitive disciplines if event log data is to be embraced as an important tool for understanding mental health and wellbeing, including mental ill health. Event log

data and mental illness are usually investigated as separate aspects within literature reviews and within single intervention studies and rarely feature in the same journals. For instance, if event log/mental illness studies are written through a more “technical” lens, then these studies are published in computational journals, whereas those event log/mental illness articles that are written through the perspective of mental illness itself would be published in psychiatry-based journals. There is a lack of studies that touch on the technicalities of the intervention, the potential of event log analysis and the theoretical basis (of which the intervention is based upon) in equal measure. This leads to a limited consensus that digital interventions work in improving mental health outcomes, which is known as the “Black Box” phenomenon (Oinas-Kukkonen. 2013; Kelders et al. 2016). This is apparent within systematic reviews when studies are compared on different interventions that the studies differ entirely based on what they report (Morrison et al. 2012). When these studies are spread across different disciplines, it hinders the development of efficient digital interventions; if a more “agile” science of event log research is to be of value across different disciplines.

4.2 Adoption of event log data by psychiatry

Typically, the information within event logs (i.e. symptom tracking information) on digital health interventions are mainly restricted to research settings and are not usually relayed back to the patients or psychiatrists to help inform treatment, which could limit the benefits that can be accrued from digital health interventions and also decrease the likelihood of care providers using these interventions as (Wallace et al, 2016). While event log data has many benefits to researchers, such as providing information about what users are doing in abundance, little is known about the reasons behind why the user is carrying out a specific behaviour or experiencing specific conditions. This information may not be obvious within log data, unless complemented by EMA. Log data cannot inform us about previous experience of the user, the intent or beliefs of the user (Dumais et al. 2014). This warrants scepticism from practitioners outside of the data analytic disciplines (e.g., psychiatry), who may require this information to base treatment plans on, plus rely on face-to-face interactions with the patient to identify cues that are salient with specific conditions through body language and speech; there may be less interaction with the digital technology than there would be in person. In terms of their efficacy in recording treatment progress, event

log data within digital health interventions (mHealth, web-based interventions) and face-to-face treatments are rarely compared to each other (Hilty et al, 2017).

4.3 Data protection issues

The log data that is available for analysis must be protected under the appropriate privacy regulations, regardless of whether the users of the technology themselves have given informed consent. The issue of informed consent depends on whether the log data contains or is combined with personal information of the user (i.e., name, address, health history) for the technology to function in the way that it is intended (Sieverink et al. 2017). The regulations behind consent to collect user data can also be different across countries and jurisdictions. For instance, the United Kingdom has been criticised for having broad legal exemptions concerning the collection of medical-related data, citing that it is in the “public’s interest” (Rumbold & Pierscionek. 2017). Many mHealth applications collect sensitive data about the user without the user’s explicit knowledge. He and colleagues (2014) examined security issues related to capturing of log data through mHealth applications, where 33% of the applications examined in the study were found to have captured sensitive information within their log files such as the user’s GPS coordinates, “followers” and disease/medication browsing history, leading to attacks on the user i.e. identity theft, invasion of privacy, physical or psychological harm (He et al. 2014). Anonymisation techniques are put in place to protect the user’s identity, which usually involves the removal of the user’s name and address. However, removal of such data can still be insufficient as one study highlighted that the combination of just 3 pieces of information could identify 87% of US residents (i.e. zip code, birth date, sex; Sweeney. 2000, Donkin et al. 2013). Risks associated with personally identifiable and/or sensitive information may not be entirely obvious to developers; extra guidance and training may help to avoid making some mistakes that are common amongst poor security practices (Dumais et al. 2014; He et al. 2014). Privacy regulations must evolve to adequately balance the user’s privacy concerns with the quality of data collected about the user.

The European Union (EU) has replaced current data protection legislation with the General Data Protection Regulation (GDPR), which comes into effect mid-2018 and is to be

adopted by all EU states. This means that researchers collecting data from users from digital interventions can only do so once users have been given the opportunity to provide consent, allowing their personal data to be used for scientific purposes (Wellcome Trust, 2016). Perhaps impacting future research, researchers and developers may need to revise current data protection standards, especially for long term epidemiological studies. This improves the integrity of the data given from users of digital interventions and services which could possibly improve the quality of the data collected; users may become confident in disclosing information without privacy concerns, allowing researchers to create a better understanding of the user's health condition and behaviour.

5. FUTURE DIRECTIONS

It may be of interest to compare other data streams to call log data (social media data, prescriptive data) to complement current knowledge about what contributes to levels of mental health and wellbeing and what may encourage usage of crisis helplines. Social media platforms i.e., Twitter and Facebook, are known to be powerful mass-media tools to help break the stigma on mental illness (Naslund et al, 2016). However, exposure to media which can be considered as negative i.e., content that promotes violence, substance abuse, racism etc can exacerbate pre-existing conditions in those who view it (Silver et al. 2013; Neria & Sullivan, 2011). The increase of social media activity can reduce face-to-face contact with a user's friends or family, increasing isolation and reducing opportunity for social contact (Miner et al. 2014). There may be a potential correlation with negative events/sentiment within social media content and the frequency and volume of calls within crisis helpline log data.

6. CONCLUSION

Analysis of user event logs shows promise to elicit understanding of digital and telephony-based interventions for mental health and wellbeing. Analysis of event logs can provide wide insight into how individuals objectively interact with a system and can overcome some of the limitations of traditional behavioural data collection methods, as mentioned earlier. Applying machine learning techniques to event log data provides researchers and practitioners with the ability to predict future mental health and wellbeing patterns and outcomes based on user

interaction behaviours, allowing for more timely-interventions and providing a glimpse of the current psychological state of a user or a population of users.

7. REFERENCES

- Al Machot, F., Mosa, A.H., Ali, M. & Kyamakya, K. 2017, "Activity Recognition in Sensor Data Streams for Active and Assisted Living Environments", *IEEE Transactions on Circuits and Systems for Video Technology*, .
- American Psychiatric Association 2014, *Desk Reference to the Diagnostic Criteria From DSM-5®*, American Psychiatric Pub.
- Armas-Cervantes, A., van Beest, N., La Rosa, M., Dumas, M. & Raboczi, S. 2017, "Incremental and interactive business process model repair in apomore", *Proceedings of the BPM Demos.CRC Press (2017, to appear)*, .
- Assent, I. 2012, "Clustering high dimensional data", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 4, pp. 340-350.
- Aung, M.H., Matthews, M. & Choudhury, T. 2017, "Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies", *Depression and anxiety*, .
- Auria, L. & Moro, R.A. 2008, "Support vector machines (SVM) as a technique for solvency analysis", .
- Bellman, R.E. 2015, *Adaptive control processes: a guided tour*, Princeton university press.
- Bradway, M., Pfuhl, G., Grøttland, A., Ribu, L. and Årsand, E., 2018, February. The Necessity of Using Mixed Methods for Assessment of mHealth Interventions: Application in the "Full Flow of Data-sharing" Project. In *Diabetes Technology & Therapeutics* (Vol. 20, pp. A43-A44).
- Bunting, B., Corry, C., O'Neill, S., Moore, A., Benson, T. & McFeeters, D. 2016, "Death by suicide: a report based on the Northern Ireland Coroner's database.", .

- Burns, M.N., Begale, M., Duffecy, J., Gergle, D., Karr, C.J., Giangrande, E. & Mohr, D.C. 2011, "Harnessing context sensing to develop a mobile intervention for depression", *Journal of medical Internet research*, vol. 13, no. 3, pp. e55.
- Cui, X., Zhu, P., Yang, X., Li, K. & Ji, C. 2014, "Optimized big data K-means clustering using MapReduce", *The Journal of Supercomputing*, vol. 70, no. 3, pp. 1249-1259.
- Dempsey, W.H., Moreno, A., Scott, C.K., Dennis, M.L., Gustafson, D.H., Murphy, S.A. & Rehg, J.M. 2017, "iSurvive: An Interpretable, Event-time Prediction Model for mHealth", *International Conference on Machine Learning*, pp. 970.
- Donkin, L., Hickie, I.B., Christensen, H., Naismith, S.L., Neal, B., Cockayne, N.L. & Glozier, N. 2013, "Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial", *Journal of medical Internet research*, vol. 15, no. 10, pp. e231.
- Dumais, S., Jeffries, R., Russell, D.M., Tang, D. & Teevan, J. 2014, "Understanding user behavior through log data and analysis" in *Ways of Knowing in HCI* Springer, , pp. 349-372.
- Eibner, C., Sturn, R. & Gresenz, C.R. 2004, "Does relative deprivation predict the need for mental health services", *J Ment Health Policy Econ*, vol. 7, no. 4, pp. 167-175.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996, "From data mining to knowledge discovery in databases", *AI magazine*, vol. 17, no. 3, pp. 37.
- Fox, S., Karnawat, K., Mydland, M., Dumais, S. & White, T. 2005, "Evaluating implicit measures to improve web search", *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 2, pp. 147-168.
- Freyne, J., Saunders, I., Brindal, E., Berkovsky, S. & Smith, G. 2012, "Factors associated with persistent participation in an online diet intervention", *CHI'12 Extended Abstracts on Human Factors in Computing Systems* ACM, , pp. 2375.
- Furukawa, T.A., Horikoshi, M., Fujita, H., Tsujino, N., Jinnin, R., Kako, Y., Ogawa, S., Sato, H., Kitagawa, N. & Shinagawa, Y. 2018, "Cognitive and Behavioral Skills Exercises Completed by Patients with Major Depression During Smartphone Cognitive Behavioral Therapy: Secondary Analysis of a Randomized Controlled Trial", *JMIR mental health*, vol. 5, no. 1, pp. e4.
- Ghahramani, Z. 2001, "An introduction to hidden Markov models and Bayesian networks", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 01, pp. 9-42.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I.H. 2009, "The WEKA data mining software: an update", *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10-18.
- Han, J., Pei, J. & Kamber, M. 2011, *Data mining: concepts and techniques*, Elsevier.
- Harvey, L., 2011. Intimate reflections: Private diaries in qualitative research. *Qualitative Research*, 11(6), pp.664-682.
- He, D., Naveed, M., Gunter, C.A. & Nahrstedt, K. 2014, "Security Concerns in Android mHealth Apps", *AMIA ...Annual Symposium proceedings*. AMIA Symposium, vol. 2014, pp. 645-654.
- Hilty, D.M., Chan, S., Hwang, T., Wong, A. & Bauer, A.M. 2017, "Advances in mobile mental health: opportunities and implications for the spectrum of e-mental health services", *mHealth*, vol. 3.
- Horvitz, E., Breese, J., Heckerman, D., Hovel, D. & Rommelse, K. 1998, "The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users", *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* Morgan Kaufmann Publishers Inc., , pp. 256.
- Incel, O.D., Kose, M. & Ersoy, C. 2013, "A review and taxonomy of activity recognition on mobile phones",

- BioNanoScience*, vol. 3, no. 2, pp. 145-171.
- Iváncsy, R. & Vajk, I. 2006, "Frequent pattern mining in web log data", *Acta Polytechnica Hungarica*, vol. 3, no. 1, pp. 77-90.
- Jain, A.K., Murty, M.N. & Flynn, P.J. 1999, "Data clustering: a review", *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264-323.
- Jakobson, G. & Weissman, M. 1995, "Real-time telecommunication network management: extending event correlation with temporal constraints" in *Integrated Network Management IV* Springer, , pp. 290-301.
- Jimoh, F., Lund, E.K., Harvey, L.J., Frost, C., Lay, W.J., Roe, M.A., Berry, R. and Finglas, P.M., 2018. Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study. *JMIR mHealth and uHealth*, 6(1), p.e17.
- Kalafat, J., Gould, M.S., Munfakh, J.L.H. & Kleinman, M. 2007, "An evaluation of crisis hotline outcomes part 1: Nonsuicidal crisis callers", *Suicide and Life-threatening behavior*, vol. 37, no. 3, pp. 322-337.
- Kelders, S.M., Oinas-Kukkonen, H., Oörni, A. & van Gemert-Pijnen, J.E. 2016, "Health behavior change support systems as a research discipline; a viewpoint", *International journal of medical informatics*, vol. 96, pp. 3-10.
- Kelders, S.M., Bohlmeijer, E.T. & Van Gemert-Pijnen, J.E. 2013, "Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial", *Journal of medical Internet research*, vol. 15, no. 8, pp. e172.
- Klemettinen, M. 1999, *A knowledge discovery methodology for telecommunication network alarm databases*, University of Helsinki Helsinki, Finland.
- Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. & Riedl, J. 1997, "GroupLens: applying collaborative filtering to Usenet news", *Communications of the ACM*, vol. 40, no. 3, pp. 77-87.
- Krishnan, N.C. & Cook, D.J. 2014, "Activity recognition on streaming sensor data", *Pervasive and mobile computing*, vol. 10, pp. 138-154.
- Lam, H., Russell, D., Tang, D. & Munzner, T. 2007, "Session viewer: Visual exploratory analysis of web session logs", *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* IEEE, , pp. 147.
- Mans, R.S., van der Aalst, Wil MP, Vanwersch, R.J. & Moleman, A.J. 2013, "Process mining in healthcare: Data challenges when answering frequently posed questions" in *Process Support and Knowledge Representation in Health Care* Springer, , pp. 140-153.
- Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S. & van der AALST, W. 2008, "Process mining techniques: an application to stroke care", *Studies in health technology and informatics*, vol. 136, pp. 573.
- Mantani, A., Kato, T., Furukawa, T.A., Horikoshi, M., Imai, H., Hiroe, T., Chino, B., Funayama, T., Yonemoto, N., Zhou, Q. & Kawanishi, N. 2017, "Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial", *Journal of medical Internet research*, vol. 19, no. 11, pp. e373.
- Maoz, M. 2013, "How IT Should Deepen Big Data Analysis to Support Customer-Centricity", *Gartner G00248980*, .
- Miner, L., Bolding, P., Hilbe, J., Goldstein, M., Hill, T., Nisbet, R., Walton, N. & Miner, G. 2014, *Practical predictive analytics and decisioning systems for medicine: Informatics accuracy and cost-effectiveness for healthcare administration and delivery including medical research*, Academic Press.
- Minor, B., Doppa, J.R. & Cook, D.J. 2015, "Data-driven activity prediction:

- Algorithms, evaluation methodology, and applications", *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*ACM, , pp. 805.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A. 2012, *Foundations of machine learning*, MIT press.
- Montgomery, D.C., Jennings, C.L. & Kulahci, M. 2015, *Introduction to time series analysis and forecasting*, John Wiley & Sons.
- Morita, M. & Shinoda, Y. 1994, "Information filtering based on user behavior analysis and best match text retrieval", *SIGIR'94*Springer, , pp. 272.
- Morrison, L.G., Yardley, L., Powell, J. & Michie, S. 2012, "What design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis", *Telemedicine and e-Health*, vol. 18, no. 2, pp. 137-144.
- Morrison, C. & Doherty, G. 2014, "Analyzing engagement in a web-based intervention platform through visualizing log-data", *Journal of medical Internet research*, vol. 16, no. 11, pp. e252.
- Moskowitz, D.S. & Young, S.N. 2006, "Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology", *Journal of psychiatry & neuroscience : JPN*, vol. 31, no. 1, pp. 13-20.
- Mulvenna, M.D., Bond, R.B., Grigorash, A., O'Neill, S., Ryan, A.A. 2018, "HILDA - A Health Interaction Log Data Analysis Workflow to Aid Understanding of Usage Patterns and Behaviours", , no. Symposium on Social Interactions in Complex Intelligent Systems (SICIS-2018), 4-6 April 2018, Liverpool, UK.
- Naslund, J., Aschbrenner, K., Marsch, L. & Bartels, S. 2016, "The future of mental health care: peer-to-peer support and social media", *Epidemiology and psychiatric sciences*, vol. 25, no. 2, pp. 113-122.
- Neria, Y. & Sullivan, G.M. 2011, "Understanding the mental health effects of indirect exposure to mass trauma through the media", *Jama*, vol. 306, no. 12, pp. 1374-1375.
- Oinas-Kukkonen, H. 2013, "A foundation for the study of behavior change support systems", *Personal and ubiquitous computing*, vol. 17, no. 6, pp. 1223-1235.
- Oliner, A., Ganapathi, A. & Xu, W. 2012, "Advances and challenges in log analysis", *Communications of the ACM*, vol. 55, no. 2, pp. 55-61.
- Olson, J.S. & Kellogg, W.A. 2014, *Ways of Knowing in HCI*, Springer.
- Quinlan, J.R. 1986, "Induction of decision trees", *Machine Learning*, vol. 1, no. 1, pp. 81-106.
- Rumbold, J.M. & Pierscionek, B. 2017, "The Effect of the General Data Protection Regulation on Medical Research", *Journal of medical Internet research*, vol. 19, no. 2, pp. e47.
- Saeb, S., Zhang, M., Karr, C.J., Schueller, S.M., Corden, M.E., Kording, K.P. & Mohr, D.C. 2015, "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study", *Journal of medical Internet research*, vol. 17, no. 7, pp. e175.
- Sieverink, F., Kelders, S., Poel, M. & van Gemert-Pijnen, L. 2017, "Opening the Black Box of Electronic Health: Collecting, Analyzing, and Interpreting Log Data", *JMIR research protocols*, vol. 6, no. 8, pp. e156.
- Silver, R.C., Holman, E.A., Andersen, J.P., Poulin, M., McIntosh, D.N. & Gil-Rivas, V. 2013, "Mental-and physical-health effects of acute exposure to media images of the September 11, 2001, attacks and the Iraq War", *Psychological Science*, vol. 24, no. 9, pp. 1623-1634.
- Spittal, M.J., Fedyszyn, I., Middleton, A., Bassilios, B., Gunn, J., Woodward, A. & Pirkis, J. 2015, "Frequent callers to crisis helplines: Who are they and why do they

- call?", *Australian & New Zealand Journal of Psychiatry*, vol. 49, no. 1, pp. 54-64.
- Steinbach, M., Ertöz, L. & Kumar, V. 2004, "The challenges of clustering high dimensional data" in *New directions in statistical physics* Springer, , pp. 273-309.
- Sweeney, L. 2000, "Simple demographics often identify people uniquely", *Health (San Francisco)*, vol. 671, pp. 1-34.
- Szigethy, E., Solano, F., Wallace, M., Perry, D.L., Morrell, L., Scott, K., Bell, M.J. & Oser, M. 2018, "A study protocol for a non-randomised comparison trial evaluating the feasibility and effectiveness of a mobile cognitive-behavioural programme with integrated coaching for anxious adults in primary care", *BMJ open*, vol. 8, no. 1, pp. e019108-2017-019108.
- Tax, N., Verenich, I., La Rosa, M. & Dumas, M. 2017, "Predictive business process monitoring with LSTM neural networks", *International Conference on Advanced Information Systems Engineering* Springer, , pp. 477.
- Thickett, J. 2006, "Connecting Older People: Consumer Engagement with Digital Services", *London, Ofcom*, .
- Tukey, J.W. 1977, *Exploratory data analysis*, Reading, Mass.
- Vaarandi, R. 2006, "Simple Event Correlator for real-time security log monitoring", *Hakin9 Magazine*, vol. 1, no. 6, pp. 28-39.
- Vaarandi, R. 2005, *Tools and Techniques for Event Log Analysis*, Tallinn University of Technology Press.
- Vaarandi, R. & Grimaila, M.R. 2012, "Security event processing with simple event correlator", *Information Systems Security Association Journal*, vol. 10, no. 8, pp. 30-37.
- Vaarandi, R. & Pihelgas, M. 2015, "LogCluster-A data clustering and pattern mining algorithm for event logs", *Network and Service Management (CNSM), 2015 11th International Conference on* IEEE, , pp. 1.
- Vaarandi, R. & Pihelgas, M. 2014, "Using security logs for collecting and reporting technical security metrics", *Military Communications Conference (MILCOM), 2014 IEEE*, , pp. 294.
- Van Gemert-Pijnen, J.E., Kelders, S.M. & Bohlmeijer, E.T. 2014, "Understanding the usage of content in a mental health intervention for depression: an analysis of log data", *Journal of medical Internet research*, vol. 16, no. 1, pp. e27.
- Walsh, S., Golden, E. & Priebe, S. 2016, "Systematic review of patients' participation in and experiences of technology-based monitoring of mental health symptoms in the community", *BMJ open*, vol. 6, no. 6, pp. e008362-2015-008362.
- Wellcome Trust 2016, July 2016-last update, *Analysis: Research and the General Data Protection Regulation-2012/0011(COD)* [Homepage of Wellcome Trust], [Online]. Available: <https://wellcome.ac.uk/sites/default/files/nw-data-protection-regulation-key-clauses-wellcome-jul16.pdf> [2018, 2018-02-13].
- Wiesner, M. & Pfeifer, D. 2014, "Health recommender systems: concepts, requirements, technical basics and challenges", *International journal of environmental research and public health*, vol. 11, no. 3, pp. 2580-2607.
- Wirth, R. & Hipp, J. 2000, "CRISP-DM: Towards a standard process model for data mining", *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* Citeseer, , pp. 29.
- Witten, I.H., Frank, E., Hall, M.A. & Pal, C.J. 2016, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.
- Zapata, B.C., Fernández-Alemán, J.L., Idri, A. & Toval, A. 2015, "Empirical studies on usability of mHealth apps: a systematic literature review", *Journal of medical systems*, vol. 39, no. 2, pp. 1.