# The non-existent average individual

Blaauw, Frank



*Publication date:*
2018

Mainly based on:

Blaauw, F. J., de Vos, S., Wanders, R. B. K., de Jonge, P., Aiello, M., Penninx, B., Wardenaar, K., Emerencia, A. C., (2017). Applying machine learning to patient self-report data for predicting adverse depression outcomes. *In preparation*.

Van der Krieke, L., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., . . . de Jonge, P. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, 25(2), 123–144.

# Chapter 2

# E-mental Health and Personalized Psychiatry

Innovations in information and communication technology (ICT) are shifting the way we deal with health care and health care delivery. The term that is indivisible from this shift is *eHealth* (Dumont, 2010). EHealth first appeared in scientific literature around the turn of the century, and is a term used to describe the use of ICT to support health care, to perform health care, or to carry out health care related research (Oh, Rizo, Enkin, & Jadad, 2005; Pagliari et al., 2005). Perspectives on eHealth (electronic health) and *mHealth* (mobile health) have changed greatly in the last decade (Fiordelli, Diviani, & Schulz, 2013; Meier, Fitzgerald, & Smith, 2013). When eHealth is applied in the field of mental health, it is often called *e-mental health* (Riper et al., 2010).

E-mental health is an umbrella term for the computer-aided practice of mental health research and practice (Riper et al., 2010; Schmidt & Wykes, 2012), or as defined by Christensen, Griffiths, and Evans (2002), "*mental health services and information delivered or enhanced through the Internet and related technologies*" (p. 17). On the one hand, e-mental health covers topics such as computer-aided psychotherapy (Marks, Cavanagh, & Gega, 2007; Proudfoot, 2004) or self-help / self-management tools (Kenwright, Liness, & Marks, 2001; van der Krieke, Wunderink, Emerencia, de Jonge, & Sytema, 2014). On the other hand, e-mental health covers service delivery (Lal & Adair, 2014), and entails mobile applications and wearable platforms to measure features related to psychopathology (Areàn, Hoa Ly, & Andersson, 2016; van der Krieke et al., 2014). In general one can think of e-mental health as the convergence between ICT solutions and mental health care.

E-mental health applications have recently gained popularity as a result of developing technologies to leverage advantages over traditional care, as illustrated using the following four points. Firstly, e-mental health applications are inherently scalable, whereas traditional care involves one-to-one relations between patient and clinician. E-mental health technology enables mental health researchers to carry out studies on a larger scale than would have been possible using traditional methods. Secondly, the use of technology can provide means for interactive and automated analysis methods. Using an automated method for analyzing data might even be inevitable in large-scale studies. As vast amounts of data are collected for ever larger groups of people, the data might grow too large for manual analysis. Additionally, manual analysis can result in inconsistent or opinionated outcomes, which can be reduced by automatizing the procedure. Thirdly, electronic data formats can facilitate interoperability in a way that medical data on paper cannot. Such data can be stored on storage devices connected to the Internet, and make ones medical data accessible world-wide. This may be of crucial importance, for instance, when a person with mental health problems faces a crisis and needs immediate help when on holiday. By having access to their medical information, the patient can immediately provide doctors with the necessary information in order to receive the right treatment. Finally, the flexibility of a Web application warrants that improvements in the care program exposed through the application will immediately benefit all applicable users.

To provide a general understanding of the fundamental concepts that underlie this dissertation, we shed light on the health care aspects and the computer science aspects of e-mental health, and on some technologies related to e-mental health. We first provide insight into the application and state of the art of precision medicine in the area of psychopathology. We then continue this trend of precision medicine and reflect on the time series methodology as currently applied in mental health research. Finally, we provide an overview of techniques currently available for analyzing such data, both from a traditional, statistical perspective, and a more recent machine learning perspective.

## 2.1 Precision Medicine

The technology component in E-mental health adds flexibility to mental health care in the sense that it can help to tailor treatments to the needs of the individual patient (Lal & Adair, 2014). In other words: ICT can assist clinicians to offer more more personalized, more precise treatment. A concept therefore closely related to both eHealth and e-mental health is the concept of personalized care, also called *precision medicine*. This concept was already introduced in 400 BC by Hippocrates, who stated the importance of *the person* in an illness, as opposed to the illness itself (Egnew, 2009).

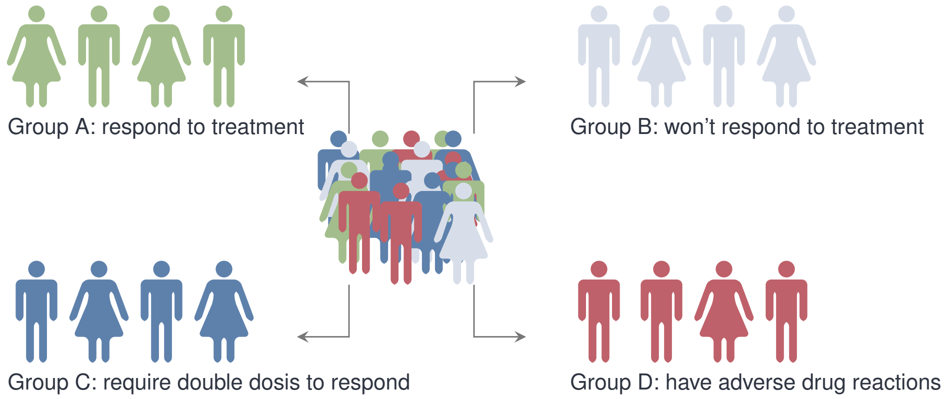



**Figure 2.1:** Precision medicine aims to provide targeted treatment plans for each group in separation, as opposed to treating the group as a whole.

> *"It is far more important to know what person the disease has than what disease the person has."*
>
> — Hippocrates, *approx. 400* BC

Personalization is defined as: *"to make personal or individual, to mark as the property of a particular person."*[1] When applied to the concept of medicine, personalization can be used to define the uniqueness of an illness on the level of the individual, instead of a general one-size-fits-all approach. It is evident that traditional doctor-patient relationships have always focused on the individual patient. Still, new advances in various fields related to medicine, such as research into the effectiveness of treatments, have only recently acted upon the importance of the person (Price, 2015). Most pharmacological research still uses clinical trials with large groups of people, which are more or less homogeneous, to test the performance of new medication, essentially focusing on 'imprecision medicine' as opposed to precision medicine (Price, 2015; Schork, 2015). This 'imprecision' medicine can be considered one of the fundamental issues regarding the shortcomings of the delivery of drugs, the infamous 'one-size-fits-all' approach (Lesko, 2007; Price, 2015; Woodcock, 2007). Figure 2.1 visualizes the differences between precision and imprecision medicine.

An example of imprecision medicine in psychiatry relates to the efficacy of antidepressants. The efficacy of antidepressants and other drugs are heavily discussed, as the effect is highly dependent on the severity of the complaints of the individ-

[1] 'Personalize.' (n.d.) In *Merriam-Webster*. Retrieved from `https://www.merriam-webster.com/dictionary/personalize`.

ual and other personal features (Fournier et al., 2010; Schork, 2015; Spear, Heath-Chiozzi, & Huff, 2001). Although antidepressants are widely prescribed to patients, their effectiveness has been reported to be as low as 50 %, while 55 % experience bothersome side effects (Bousman et al., 2017; Papakostas, 2009). Simply stated: antidepressants work for some people, but not for all. In order to improve health care, we need to know for whom it works, to what extent, and whether the effects are beneficial or detrimental for that very person. Not surprisingly, researchers have stated that precision medicine is "*the logical next step in progressing medical science*", and it is not a question of *if*, but rather a question of *when* a more personalized approach will be adopted (Woodcock, 2007).

The inevitable rise of precision medicine is — and will be — largely fueled by advances in information technology and advances in health care research (Downing, Boyle, Brinner, & Osheroff, 2009; Lesko, 2007). Technological advances in health care research and in general society allowed for great strides in precision medicine over the last decades (Louca, 2012). One of the most important advances that can fuel a more precise medicine can be considered the revolution in DNA sequencing. Advances in genome technology and the use of ICT in the analysis of DNA resulted in massive price drops of the analytical process (Check Hayden, 2014; Ozomaro, Nemeroff, & Wahlestedt, 2013), making the technique available to a wide public. Nowadays, consumers can order their personal DNA research, in which their saliva is tested for various health risks, for 149 dollars[2]. Other technological advances have been made in the field of big data analysis. For example, researchers have shown how big data analysis could be used to detect adverse drug interactions, by collecting and analyzing search engine queries (White, Tatonetti, Shah, Altman, & Horvitz, 2013). The ability to quickly and accurately analyze large datasets enables the applicability of precision medicine (Panahiazar, Taslimitehrani, Jadhav, & Pathak, 2014; Swan, 2012a; Viceconti, Hunter, & Hose, 2015).

There are a few examples of successful implementations of precision medicine. For instance, in oncology, treatments have been personalized in some cases by extending traditional diagnostic measures with various gene expression-based measures in order to improve treatment decisions (Kalia, 2015; Mehta, Jain, & Badve, 2011). In (psycho)pharmacology, researchers found out that the response to a drug can be influenced by various personal factors (such as genetics, demographics, environmental factors, etc.; Crisafulli et al., 2011; Hamburg & Collins, 2010; Mancinelli, Cronin, & Sadée, 2000; Mroziewicz & Tyndale, 2010; Tansey et al., 2013). Researchers predict that the effectiveness of antidepressants could be increased by including genetic profiles of the individual that guide the prescription of the dosage of drugs to individuals drug (Katsanis, Javitt, & Hudson, 2008).

[2]See for example `https://23andme.com`.

Although successful examples exit, precision medicine is still in its infancy (Lesko, 2007; Personalized Medicine Coalition, 2014). As Lesko (2007) puts it: "*Personalized medicine is a paradigm that exists more in conceptual terms than in reality*" (p. 807). This is the case for general medicine, but it also holds for the area of psychiatry and psychopathology (Ozomaro, Wahlestedt, & Nemeroff, 2013). The following sections will conceptualize precision medicine in psychiatry and also glimpse at reality. In Section 2.2 we describe how time series data can be used to offer insights into psychopathology in a personalized fashion. Section 2.3 further elaborates on precision medicine, and describes various analysis methods that can be performed on (time series) data collected in psychopathology research.

## 2.2 Psychopathology as a Time Series

Traditionally, research in psychopathology is predominantly based on nomothetic studies, that is, the process of unveiling statistical parameters from large population studies (van der Krieke, 2014). Such research has the advantage of being straightforward to conduct; recording a single measurement per person in a large enough population sample can be easily done with todays technology. However, this simplicity comes at a price. The disadvantage of nomothetic research is that the outcome is based on population averages, to which the individual is generalized. For instance, if research into the effectiveness of antidepressants shows that they are effective, than the assumption is that it will be effective for each depressive individual. This generalization step does not always provide sensible, correct results and might lead to the ecological fallacy (Piantadosi et al., 1988). Instead of focusing on the population as a whole, one could also shift focus to the individual, in a more idiographic way (Allport, 1937). Idiographic research focuses on *intra*individual variability, instead of the variation of the group; *inter*individual variability. In idiographic research the individual would preferably be measured multiple times over a certain period of time, so one can grasp the moment-to-moment variability within each individual (Diggle, Heagerty, Liang, & Zeger, 2013).

Psychopathology research is rooted in methods that work with data retrieved from validated questionnaires and self-reports (Danziger, 1990). While most physical ailments have objectively measurable symptoms, mental disorders are typically more subjective, and clinical practice has come to rely on establishing the presence of psychopathology through sets of validated questionnaires and interviews. A recent trend in psychopathology research is the use of intensive longitudinal studies, in which people monitor their mental health frequently (e.g., daily), for a longer period of time using *diary studies*. In Section 1.2 we introduced two well-known methods for collecting intensive self-report data; the experience sampling method (ESM; Csikszentmihalyi & Larson, 1987) and the ecological momentary assessment (EMA; Shiffman, S., & Stone, 1998). These methods have been successfully applied in various studies on psychopathology, for instance studies focusing on stress and depression (Booij et al., 2015), mindfulness and depression (Snippe et al., 2015), pain (Stone et al., 2003), mood disorders such as major depressive disorder (MDD) and bipolar disorders (aan het Rot, Hogenelst, & Schoevers, 2012; Ebner-Priemer & Trull, 2009), and melatonin secretion and depression (Bouwmans et al., 2015). By measuring a person intensively, in their natural context, and for a longer period of time, moment-to-moment changes in experiences, psychological factors, context, and behavior can be mapped out. For example, at moment one, a person might rate their level of happiness a five (e.g., on a scale from one to ten), while at moment two, their happiness could have increased to an eight. Apart from measuring this intraindividual variation, collecting data using diary studies has several other advantages over traditional, cross-sectional methods. Diary studies allow for measuring people in their natural context, and as such capture not just features regarding the individual, but also the interaction of the individual with the environment or current context (Reis, 1994). Furthermore, collecting data in an intensive longitudinal study reduces the effect of recall bias, or retrospection, and can thus increase data reliability (Bolger, Davis, & Rafaeli, 2003). Recall bias (deviations from the truth caused by inaccurate recollections of past events and experiences) is generally reduced as the time between events and measurements is decreased (Solhan, Trull, Jahng, & Wood, 2009).

Collecting and processing such diary study data was initially a challenging and cumbersome task (Trull & Ebner-Priemer, 2009). Early studies that used such diary studies applied methods that required pencil and paper to collect measurements (e.g., Wichers et al., 2007). Collecting data using such analogue means has several disadvantages. The first obvious disadvantage is that this general procedure is tedious, especially for transferring the data when the participants are highly distributed nationally, or even globally. Second, from a methodological perspective, such method could be considered less valid. For EMA it is important that participants fill out the questionnaires at fixed (or predefined random, depending on the study protocol), chronological moments. Issues could present themselves if the responsibility of filling out these questionnaires lies at the individual, such as forgetting to fill out a measurement, or 'back-filling' earlier missed measurements (Trull & Ebner-Priemer, 2009).

Fortunately, nowadays with the wealth and ubiquity of ICT, conducting large scale idiographic studies is relatively easy. Research for which a large number of assessments need to be conducted (possibly multiple times per day) can now be performed digitally using mobile technology. Specialized devices exist to enable collecting regular self-reports (e.g., the dedicated *PsyMate*[3]). Instead of using specialized diary devices for conducting self-measurements, one could take advantage of the popularity of smartphones to perform such studies (Jebb et al., 2015). It is evident that the modern individual possesses more mobile technology and 'smart' devices than ever before (Spencer Trask and Co, 2014). Some of these devices, such as smartphones and tablets, enable people to have Internet access during a large part of their daily life, allowing the use of new and more accurate methods to perform assessments in health care and health research (Trull & Ebner-Priemer, 2009). As such, various generic mobile applications exist to perform EMA, such as LocasaESM (Patil & Lee, 2013), Movisens (Movisens GmbH, 2017), mEMA (Ilumivu, 2015), ULTEMAT (van de Ven et al., 2017), and Ohmage (Ramanathan, 2012), to name a few. In Chapter 3 we describe two of our own platforms for collecting EMA data.

The ease with which personal data can be collected has given rise to a data collection movement outside the area of scientific research — the so-called the Quantified Self (QS) movement (Swan, 2012a, 2012b; Wolf, Carmichael, & Kelly, 2010). Swan (2009) defines QS as: "*the regular collection of any data that can be measured about the self such as biological, physical, behavioral or environmental information*" (p. 509). The popularity of QS has resulted in the development of a wide variety of health sensors and wearables to measure the 'objective' parameters of people their health, and online platforms to support their use. Well-known examples are Fitbit[4] and Google Fit[5] Swan (an elaborate list of techniques, wearable devices, and applications is provided in 2012b, and in the QS guide[6]).

A next step in the personalization of psychopathology research is to combine intensive idiographic self-report data with the more objectively measured data of sensors and wearables. A few small scaled studies have experimented with this combination. For example, Booij et al. (2015) performed an EMA study wherein the participants wore an accelerometer to measure physical activity whilst filling out EMA questionnaires about mood, cognition, and daily activities. Furthermore, M. Schenk (2017) researched the interplay between physiological biomarkers and mental health. However, studies like these often use expensive, single-purpose devices for the sensor measurements. The cost of these devices impedes large scale integration of sensor technology in EMA studies. With the increasing popularity and quality of smartwatches and other sensor-equipped wearables in recent years, it becomes possible to use sensor data from wearables that a participant is already wearing. Integrating these commercially available wearables for use in large scale

[3]Website: `http://psymate.eu`.
[4]Website: `http://fitbit.com`.
[5]Website: `https://google.com/fit/`.
[6]Website: `http://quantifiedself.com/guide`.

EMA studies used to be a non-trivial task due to the diversity of the different service providers and the incompatibility of the exported data formats with EMA data. However, in Chapter 9 we describe our Physiqual platform, a platform that relieves such challenges and provides the functionality to automatically merge such data.

## 2.3 Predicting and Explaining Psychopathology

Raw data collected using the EMA methodology or wearables can easily be used to generate simple visual representations and overviews, for instance of the course of someone's mood and physical activity from day to day. However, in order to generate information about interactions between variables, about how physical activity predicts mood or the other way around, when, and to what extent, advanced statistical analysis is needed. An often used method for the analysis of intensive idiographic studies in psychopathology is time series analysis. The next section will provide an overview of this method. After that, at the end of the chapter, machine learning will be introduced as an alternative analysis method.

### 2.3.1 Time Series Analysis

Time series analysis is a methodology useful in a variety of fields apart from psychopathology research, for instance econometrics (e.g., Granger, Newbold, & Shell, 1986), meteorology (e.g., Duchon & Hale, 2012), power systems (e.g., Hagan & Behr, 1987), and astronomy (e.g., Vio, Kristensen, Madsen, & Wamsteker, 2004). As such, multiple statistical modeling techniques exist for analyzing time series data, most of which aim to construct models to investigate hypothetical events, and / or explain the underlying systems (Box & Jenkins, 1976; Lütkepohl, 2005). Basic techniques include the use of (auto)correlation to see how much certain variables in the data (e.g., measures of depression and physical activity) are related to each other. Slightly more complex, but based on the same principles, are techniques which assume statistical models such as autoregression (AR) or vector autoregression (VAR) models (e.g., Brandt & Williams, 2007; Lütkepohl, 2005). In AR, a single variable is regressed on historical values of itself. For example, depression at time $t$ could be hypothesized to be a linear function of depression at times $t-1, \ldots, t-p$, such that

$$\text{depression}(t) = \beta_0 + \beta_1 \cdot \text{depression}(t-1) + \ldots + \beta_p \cdot \text{depression}(t-p) + \epsilon,$$

where $\beta_0$ represents an intercept term, $\epsilon$ denotes the error term, and $\beta_1, \ldots, \beta_p$ the amount with which depression at time $t-1$ influences depression at the current moment in time. Instead of focusing on a univariate time series, a VAR model includes a vector of variables. VAR is a technique that originates from econometrics (Sargent, 1979; Sims, 1980), and can be used to fit a multivariate regression model; a model in which the outcome of one variable is regressed on the outcomes of several other variables, while including an autoregressive component. In VAR there is no clear distinction between input and output (viz., dependent and independent variables), and each variable is considered both an input and output variable. Using the previous example and augmenting it with an 'activity' variable gives us

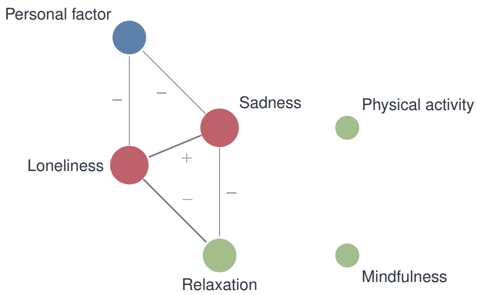
$$\text{depression}(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \cdot \text{depression}(t-1) + \sum_{i=1}^{p} \beta_i \cdot \text{activity}(t-1) + \epsilon_1,$$
$$\text{activity}(t) = \beta_0 + \sum_{i=1}^{p} \gamma_i \cdot \text{depression}(t-1) + \sum_{i=1}^{p} \delta_i \cdot \text{activity}(t-1) + \epsilon_2,$$

in which the $\alpha_i$, $\beta_i$, $\gamma_i$, and $\delta_i$ coefficients show the amount with which previous values at time $t-i$ influence a variable's current value, and $p$ denotes the number of lags, or historical time points to consider (a lag is the difference in time between an observation and a previous observation). Other statistical models include moving average (MA), autoregressive moving average (ARMA), or autoregressive integrated moving average (ARIMA) models (e.g., Brandt & Williams, 2007; Jebb et al., 2015), or autoregressive conditional heteroskedasticity (ARCH) processes (Engle, 1982). In such MA models, the current value of an outcome is not only regressed on previous values of the modeled variables, but also on the previous error terms, such that

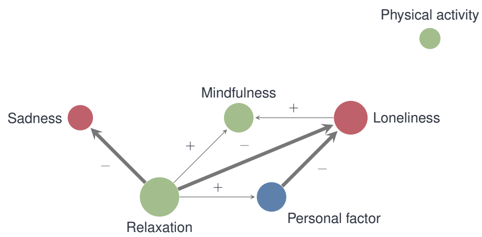
$$\text{depression}(t) = c + \epsilon(t) + \sum_{i=1}^{p} \beta_i \cdot \text{depression}(t-i) + \sum_{i=1}^{q} \theta_i \cdot \epsilon(t-i),$$

in which the various coefficients are the similar to the ones AR and VAR models, $q$ is the number of moving average terms, $\theta_i$ the influence of that error term on the present measurement, and $c$ is a constant (e.g., the mean). These methods are all based on regressing a set of variables on another (or the same) set of variables, in order to retrieve a set of regression coefficients. These regression coefficients are then used to investigate the relations between modeled variables. Applied to psychopathology research, an example of a research question could be: is a decrease in depression predicted by an increase in physical activity, and does a decrease in depression on moment one predict a further decrease on moment two? A general overview of different time series analysis and forecasting techniques is provided by Box and Jenkins (1976); Chatfield (1996); and Lütkepohl (2005).

As AR (and other time series analysis techniques) take the notion of time into account, they can be used to introduce notions of causality, such as *Granger causality* (Granger, 1969). Granger causality is a notion of causality describing that the variance of one variable ($x$) can better be explained using time lagged values of both $x$ and of another variable ($y$) instead of merely using lagged values of $x$. In such cases, $y$ is said to Granger cause $x$ (Granger, 1969). Insight into these Granger causalitys can be given by means of *network models* which depict the relations of these variables, as depicted in Figure 2.2. In these network representations, we use red nodes to represent variables that tend to be perceived as negative (sadness, loneliness). We use green nodes to represent variables that tend to be perceived as positive (relaxation, mindfulness, physical activity). We use the blue node to represent a variable of which the interpretation is unknown[7]. The size of the node indicates its relative importance (i.e., the bigger the node, the more relationships that variable has with other variables; also known as 'degree'). The lines represent the relationship between variables; the thickness indicates the strength of the relationship. A plus refers to a positive relationship; a minus refers to a negative relationship. The arrowheads (only in the dynamic network) indicate the direction of the relationships.




**(a)** Contemporaneous network.




**(b)** Dynamic network.

**Figure 2.2:** Network images showing the relation between the six modeled variables.

The first network model (Figure 2.2a) depicts the contemporaneous (or concurrent) relationships between variables. This model shows how a participant's affect, cognitions, and behaviors are related to each other at the same moment in time. This contemporaneous network shows that (i) when this person felt lonely, this person was more sad, less relaxed, and experienced less of the personal factor; (ii) when this person was more relaxed, this person was less sad (and felt more lonely); and (iii) when this person felt more sad, this person was less relaxed, more lonely, and experienced less of the personal factor. Physical activity and mindfulness did not show any contemporaneous relations. Furthermore, the size of the nodes indicate

[7]In the study that collected the data for these network images participants were allowed to select one variable or symptom themselves.

that loneliness and sadness had most connections with other variables (also known as degree), and the thickness of the line indicates that the relationship between loneliness and sadness, and loneliness and relaxation were the strongest relationships. The relationships in the contemporaneous network show that an individual their mood, emotions, and behaviors are related to each to other at the same moment in time, but do not indicate the temporal ordering of these associations. The second network model (Figure 2.2b) illustrates the temporal ordering of the relationships between variables (namely, how variables affect each other across time). The dynamic network for this person shows that (i) an increase in relaxation preceded (thus predicted) a decrease in sadness, an increase in mindfulness, a decrease in loneliness, and an increase of the personal factor; (ii) an increase in loneliness predicted an increase in mindfulness; (iii) an increase in the personal factor predicted a decrease in loneliness; and (iv) physical activity did not predict, nor was it predicted by, any of the included variables.

Instead of analyzing the coefficients and Granger causality of such statistical models, another way to investigate these relations is by using simulations, forecasts, or hypothetical intervention techniques (Jebb et al., 2015). Such techniques could provide insight into the situation, and how the model would react to certain changes (Borsboom et al., 2016). Simulations can be used in order to see what a model could do over time, and how the variables in the model react to changes of each other. Various techniques exist for doing simulations, however, some of them are specifically useful for VAR models. Some of these methods include impulse response function (IRF; Brandt & Williams, 2007; Lütkepohl, 2005), orthogonalized impulse response function (OIRF; Brandt & Williams, 2007), Monte-Carlo simulation (Zivot & Wang, 2006), or more related to machine learning techniques, (recurrent) neural network forecasting (Dorffner, 1996; Zhang, Eddy Patuwo, & Hu, 1998). We mainly focus on methods of simulation that can be performed on VAR models, specifically IRF (Blaauw, van der Krieke, Emerencia, Aiello, & de Jonge, 2017b; Brandt & Williams, 2007). In IRF, a sudden increase (or shock) is imposed on one of the variables and it is observed how the system responds to such intervention. Applied to psychopathology, an example could be: physical exercise is employed by a depressive person and that exercise might affect the depressive symptoms for a certain period of time. In Chapter 6 we focus on a method to automatically generate information about the interaction of mental health variables by applying IRF analysis on VAR models.

Although these traditional, statistical methods are useful in many cases, they have their downsides. For instance, VAR models make several important and restrictive assumptions, such as the assumption of a linear and parametric statistical model. Such issues have given rise to the development of new, more flexible methods and perspectives in the field of statistics, computer science, and mathematics, known as machine learning. The field of machine learning offers various techniques that can be arguably more flexible and rely on less assumptions than traditional techniques. This machine learning perspective is described in the next section.

### 2.3.2 Machine Learning Perspective

Largely fueled by the increasing availability of computer resources, such as large storage devices, ever faster central processing units, and the widespread availability of fast graphical processing units, the use of and interest in machine learning methods is greater than ever before (Jordan & Mitchell, 2015). This popularity is exemplified by the normalized number of papers published on *ArXiv* with respect to machine learning or the Google search and news trends related to machine learning in Figure 2.3[8] (Google Inc., 2017).

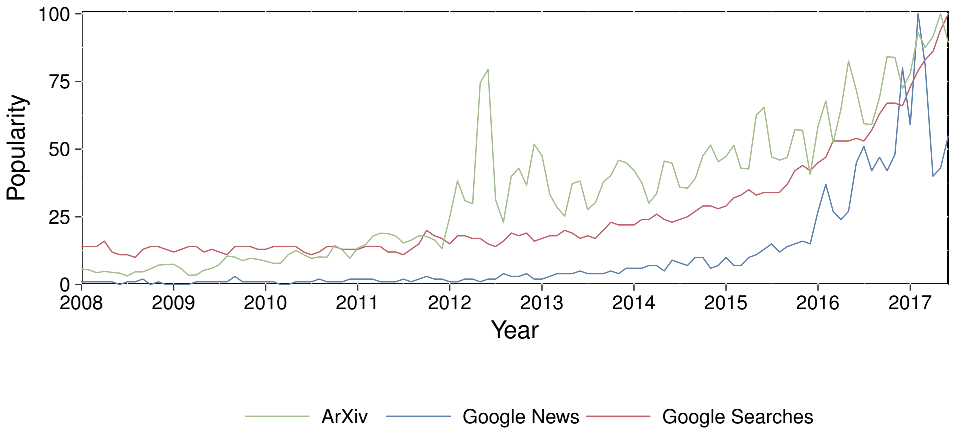



**Figure 2.3:** Popularity of machine learning based on the normalized ArXiv publications in the categories cs.AI, cs.LG, cs.CV, cs.CL, cs.NE, and stat.ML (categories retrieved from Karpathy, 2017), number of Google Web-searches for machine learning, and Google News results for and related to the term machine learning over time (Google Inc., 2017). Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise a score of 0 means the term was less than 1 % as popular as the peak.

[8]Source available at `https://github.com/frbl/thesis-supplement`.

We define machine learning as the process that enables an algorithm, or an ensemble of algorithms, to 'learn' an underlying and generally unknown mechanism from which our observations (the data set) originate. As such, after an algorithm has learned how this mechanism behaves, the algorithm should be well-equipped to make predictions based on new data. Machine learning algorithms are currently available in many shapes and sizes. We consider any algorithm that could be used to create a prediction and is able to learn from data as a machine learning algorithm (e.g., [logistic] linear regressions, support vector machines, artificial neural networks, and many more; Hastie, Tibshirani, & Friedman, 2009). The goal of machine learning is similar to 'traditional' predictive statistical methods such as time series analysis, that is, to learn from data, make predictions, and find patterns in data (Cleophas & Zwinderman, 2014). Although learning or training in the traditional sense of the word is not something a machine (viz., a computer) can do, in machine learning parlance these terms are used in a similar way. For both terms the distinction can be made between *knowledge acquisition* and *skill refinement*, and both underly the general procedure of machine learning (see e.g., Michalski, Carbonell, & Mitchell, 1983, for an comparison of different machine learning techniques).

After a short standstill in the 1980s, in which the field of machine learning could not live up its expectations (Nilsson, 2009), it regained popularity from the 1990s onwards due to valuable applications in the commercial setting. Nowadays, many companies use solutions based on machine learning to offer customers with a better experience. For example, online streaming video provider *Netflix* and online retailer *Amazon* use collaborative filtering approaches to give you recommendations for a next movie to watch or product to buy (Carrer-Neto, Hernández-Alcaraz, Valencia-García, & García-Sánchez, 2012), and the social network *Facebook* uses machine learning to do face identification in images (Taigman, Yang, Ranzato, & Wolf, 2015). Machine learning also became slightly famous, when a machine learning based computer won the American game show *Jeopardy!* (Markoff, 2011).

Probably one of the most successful (and hyped) machine learning approaches currently is *deep learning*. The popularity of deep learning is mainly fostered by its impressive achievements (Hof, 2013; Jones, 2014; LeCun, Bengio, & Hinton, 2015). For example, deep learning has been used to automatically classify objects (e.g., people) in images and to automatically generate a highly accurate captions (e.g., the caption "*two young girls are playing with Lego toy*" was generated by a deep learning algorithm; Karpathy & Fei-Fei, 2015, p. 3134), a task that is rather complex for a machine to perform. Slightly overshadowed by this interest in deep learning are numerous other algorithms and platforms supporting the widespread implementation of *data science*. Data science is fundamentally the process of extracting knowledge and information from data (Provost & Fawcett, 2013). It is an interdisciplinary field that combines components from computer science, statistical inference, and mathematics to discover knowledge in data. The concept of machine learning is intrinsically linked to data science. As an example, Google's *Kaggle* platform hosts several data science competitions and offers monetary rewards for performing machine learning research (Kaggle Inc, 2016).

The value machine learning could add to medicine is tremendous, as was already demonstrated by the IBM Watson using its *DeepQA* architecture (Ferrucci, 2012). The IBM Watson applies different machine learning techniques to provide advice in, for example, medical decision making, and has already been applied successfully in several cases (Doyle-Lindrud, 2015; Knight, 2015; Murdoch & Detsky, 2013). Watson could eventually even exceed human performance and provide advice to clinicians (Steadman, 2013).

In the field of psychopathology research, the use of machine learning has also gained traction over the last years. Table 2.1 presents a strand of literature in which machine learning methods have been applied in psychopathology research. The applications of machine learning in psychopathology research are diverse. For instance, Askland et al. (2015) used machine learning to predict remission in people suffering from obsessive compulsive disorder (OCD). Machine learning has also been used to predict treatment response. For example, Amminger et al. (2015) used machine learning to investigate treatment response of omega-3 fatty acids in people with high risk for psychosis. Various other researchers focused on the prediction of effects of (combinations of) antidepressants (Chekroud et al., 2016; Dodd et al., 2014; Nelson et al., 2012; Perlis, 2013).

How these computerized methods can be applied is nicely demonstrated by a rudimentary Web application as created by Perlis (2013), in which users can input some of their personal data (for example, gender, age, and scores on several questionnaires), and get a personalized estimation for the response to a certain treatment.

The application in of machine learning in clinical research is considered the inevitable road to the future (Kononenko, 2001; Murdoch & Detsky, 2013), but its progress goes slowly due to various factors, such as privacy (Murdoch & Detsky, 2013; Schizas, 2015), limited interoperability of electronic health record (EHR) platforms (Murdoch & Detsky, 2013; Schiza, Neokleous, Petkov, & Schizas, 2015) limited interpretability of the results and decision process (Kuhn & Johnson, 2013), or general juridical and political issues (Price, 2015; Safran et al., 2007). Other factors might be the reduced mathematical rigor of machine learning approaches (Cleophas & Zwinderman, 2014), or the fact that machine learning might introduce yet another application for physicians to work with (Kononenko, 2001). In Chapters 7 and 8 we present our view on and approach for analyzing psychopathology data using a machine learning approach.

**Table 2.1:** Several studies in psychiatry that have applied machine learning methods.

| **Reference** | **Machine learning algorithms** | **Objective** | $n$ |
|---|---|---|---|
| Amminger et al. (2015) | Univariate regression, Gaussian process classification | Treatment response in young people ultra-high risk for psychosis | 81 |
| Andreescu et al. (2008) | Univariate logistic tests (for feature selection), Decision trees | Treatment response for depression | 461 |
| Askland et al. (2015) | Random forests | Identify predictors of remission in OCD | 296 |
| Chekroud et al. (2016) | Elastic net (for feature selection), Gradient boosting machine | Treatment response for depression | 4 192[a] |
| Dodd et al. (2014) | Gradient boosted model | Differences in treatment response of different antidepressants | 4 987[b] |
| Etkin et al. (2015) | Pattern classification | Use of biomarkers forremission of andidepressants | 1 008[c] |
| Galatzer-Levy, Karstoft, Statnikov, and Shalev (2014) | Support vector machines, Generalized local learning, Random forests, AdaBoost, Kernel ridge regression, Bayesian binary regression | Early forecasting of post-traumatic stress disorder (PTSD) | 957 |
| Jain, Hunter, Brooks, and Leuchter (2013) | QROC (uses decision trees) | Antidepressant response and remission | 2 876 |
| Kessler et al. (2015) | Regression trees, Elastic net penalized survival models, Discrete time survival model, Penalized regression (for feature selection) | Suicide prediction in US Army Soldiers | 53 769 |
| Nelson et al. (2012) | Logistic regression, Classification and regression trees | Identification of predictors of remission with placebo treatment in MDD | 1 017 |
| Perlis (2013) | Logistic regression, Naïve Bayes, support vector machines, random forest | Risk stratification for predicting treatment resistance in MDD | 2 555 |
| Riedel et al. (2011) | Logistic regression, Classification and regression trees, Univariate tests (for feature selection) | Identify variables for prediction of response and remission in treated inpatiants with MDD | 1 014 |
| Wall, Dally, Luyster, Jung, and DeLuca (2012) | Various tree based approaches, Incremental Reduced Error Pruning, Nearest Neighbor algorithms, Various Association rule techniques | Screening and diagnosis of autism | 627 |
| Wall, Kosmicki, DeLuca, Harstad, and Fusaro (2012) | Various tree based approaches, Incremental Reduced Error Pruning, Nearest Neighbor algorithms, Various Association rule techniques | Diagnosis of autism | 2 942[d] |

*Note:*
[a] Of which 1 949 participants compleded the study, and 425 participants were used from a separate study as test set.
[b] Pooled accross 12 ranomized controlled trials (RCTs).
[c] Of which 665 participants completed the study.
[d] Trained on 891 individuals diagnosed with autism and 75 that did not meet diagnostic criteria, tested on 1 976 individuals from two different sources.

# Part I

# Monitoring and Measuring Psychopathology Online